Understanding Transcription through Structural Characterization of the CBP-p53 Transcriptional Coactivator-Activator Complex

by

Melody Sanders

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2021

Doctoral Committee:

Professor Anna Mapp, Co-Chair
Professor Melanie Ohi, Co-Chair
Associate Professor Amanda Garner
Professor Emily Scott
Professor Janet Smith

This Dissertation Was Written By:

Melody Sanders

Eloquent Like Baldwin.

Bold Like X.

Kind Like King.

More Brilliant than the Stars.

You Will Never Meet Another Scientist with a Mind Quite Like Mine.

-Mel Sanders.

Artist. Activist.Scholar.

Melody Sanders

melsande@umich.edu

ORCID iD: 0000-0002-5654-8484

DEDICATION

The Dissertation herein is Dedicated to all Black People in this Nation. Being BLACK is Not a Barrier for You. White Supremacy Is. Stand Together United and Name Your Oppressors. I Hear You. I See You. I Support You.

-Mel

Acknowledgements

It gives me great pleasure in expressing my gratitude to all the people who have supported me and had their contribution in making my doctoral dissertation possible. However, any attempt to list the people and opportunities to which my life has been so richly blessed would be like trying to count the stars in the heavens. So, for anyone that I have unintentionally left out, I ask your forgiveness.

I will commence my acknowledgments by thanking my thesis advisors Dr. Anna Mapp and Dr. Melanie Ohi. My experiences in each of your laboratories provided me with invaluable training opportunities and have served in part as the inspiration behind much of the science and activism I do today. Thank you for giving me the autonomy to pursue my own research ideas and the space to develop into a leader and an independent scientist. By the same token, I must acknowledge the members of my dissertation committee, Dr. Amanda Garner, Dr. Janet Smith, and Dr. Emily Scott. There is not one dissertation committee as remarkable as mine. Thank you for the insightful scientific discussions.

I also have to thank members of the Mapp laboratory and the Ohi Laboratory. Specifically, I must acknowledge Dr. Clarissa Durie, Dr. Jason Porta, and Dr. Matthew Henley. Clarissa, you have been one colleague who has not only assisted me technically with my scientific endeavors but also someone who has been supportive of and present for nearly all of my advocacy campaigns. Jason, I find you brilliant beyond measure. Thank you for accepting me and being on call during my arduous protein preps in the early stages of my cryo-EM training when I was just beginning to climb the steep slopes of the cryo-EM learning curve.

Matthew, I must express my profound sense of reverence and gratitude to you. Your ability to develop and systematically approach compelling research problems, your high scientific standards, and your passion for science has always served as an inspiration to me and has been indispensable in my growth not only as a scientist but also as an individual. You have been an encouraging peer and colleague in times of my new scientific ideas and career difficulties.

To all my mentors throughout the years; Dr. Roy Welch, Dr. Gina Lee-Glauser, Dr. Ramesh Raina, Dr. Ethriam Brammer, and Dr. Cherie Dotson-the work presented here would not have been made possible without you. Each of you gave me something that many others wouldn't, a fair opportunity. Your continuous guidance, support, and encouragement have truly been instrumental in my professional development and overall success. Thank you doesn't seem sufficient for all you have done and continue to do but it is said with the utmost appreciation, respect, and sincerest gratitude.

For my friends, Ulla, Courtney, Aileen, Otto, Pampi, Garrett, Juniar, Jeff, Alys, Taylor, and Natalie- thank you for giving me community, showing me I belong, and providing me with words of encouragement during the times in which I felt like I couldn't carry on. Dr. Omari Baruti, I would be remiss if I did not make a particular mention of you. You helped me recognize my true worth, citing my excellence in spite of what others said, and always reassured me that truth, knowledge, understanding, and wisdom were the only drums that I needed to ever march to. Thank you very much for being a part of my educational journey to the doctorate and supporting me in all my social justice pursuits when many others turned away.

Dr. Ginger Shultz, you are and will remain my best role model for a scientist, mentor, colleague, and friend. When you became a part of my doctoral journey here at the University of Michigan the timing could not have been better. I don't think I could have continued to be an effective learner, diligent scientist, or even muster the courage to continue to pursue activism if I were not nourished by your constant and continual source of support. You taught me self-pride, the value of my voice, and have created one-of-a-kind opportunities for me to advance the principles of social justice while I simultaneously pursued studies for my degree. This level of active support has been instrumental in my overall success here at the University of Michigan and I am forever grateful. Thank you

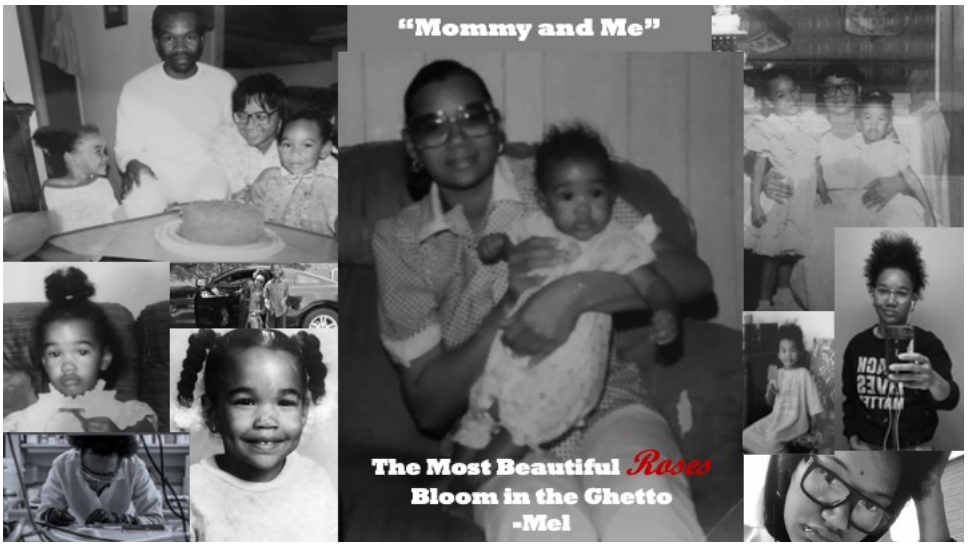
for being an outstanding leader, colleague, mentor, and ally. If we are ever to overcome the many divisions that prevent a broad-based movement for social justice in higher education, this movement will most likely be inspired and led by yourself and others like you.

Finally, I would like to send precious and warm sentiments to the members of my family. As I reflect on the years that have gone and passed, I cannot imagine where I would be in the absence of your love, guidance, patience, and support. So much of what is good in me has come from each of you. You all have been a source of wisdom and inspiration for me and a safe haven from the storm teaching me the ways of life, giving me the strength to face its challenges, and the courage to carry on. You are my family; woven together not all by genetic material but by the Grace of God and holding the most special place in my heart. Mommy and Beza, I owe the most beautiful and eloquent thanks to you, so the following pages of this acknowledgment section have captured my special sentiments for you.

For Mom

(The One That Gave Me Life)

"The Most Beautiful Roses Bloom in the Ghetto"



Dear Mama,

Words fail me when I try to express just exactly how I feel;

so I will start by saying I love you, Mama, back then, now, and I always will.

You gave me life, nurtured me, taught me, and fought for me.

Importantly, when the rest of the world turn their backs on me you loved me unconditionally.

You did your best to provide for me and spoil me with designer clothes even though we were poor.

I thank you for all these things, but I am most grateful for your precious gift of love because that to me meant so much more.

I still remember the days you brought me Jordans every time I brought home them A's.

And when I lost you, Mama, I still remember my heart breaking in a hundred thousand different ways.

You brightened my darkest days like the sun.

So, no matter how old I become, I will always remain your little one.

Before you left this world and to this very day;

I've spent my whole life trying to provide for us, trying to make a way.

At 12-years old I started hustling on the streets.

I wanted to make sure you, my sister, and I had a place to lay our heads at night and enough food to eat.

I hustled because I had to, not because it was cool.

I still made sure I brought straight A's home to you Mommy; I never missed a beat in school.

Generational cycles of poverty, drugs, violence, and crime with no education I've been steadily trying to break; all the while fully recognizing from an early age that with one single wrong turn it was my very own life that was at stake.

Even back then while hustling on the corners, Mom I had a desire for and visions of so much more.

So, I worked hard to get up out the streets and away from a lifestyle that anyone would abhor.

Driven by passion, creativity, unmatched ambition, my aspiration was to be the elevation of the next generation.

I applied to college and went on to pursue my higher education.

Looking forward to a world of new and exciting opportunities, I rushed to enter the doors of the ivory tower to make my escape from a place where my life was never promised to me.

Much to my surprise Mommy, you would never know all the hate your baby girl was going to see.

I witnessed and experienced an institutional framework that was nothing like I imagined it to be; a full manifestation of the ugliness intricately woven into the fabric of our society.

I watched grace be given to whites in power, while I broke my back to get things done.

There was no kindness, only injustice, and acts of violence that tried to break the wings of me, a young, naïve, and unsuspecting black woman.

I felt the hate and saw the expressions of racism in the academy through many blatant facts and all the insincere “we value diversity” acts.

You see, the notions of equality of opportunity in higher education weren’t set in practice only in statute,

In other words, despite the well-conceived rhetoric, opportunity was absent for those who look like me and instead was reserved and given to only a select, privileged few.

Not to mention, the young, black, and gifted kid from the ghetto, a scientist and an activist, never too tempered in speaking the truth, became the paradox that didn’t quite fit into the institute's merit box.

The science game wasn’t easy, but I mastered it, got some experience in the industry, and then got my doctorate.

I have a great story, from the ghetto to glory, so I put it in print.

I captured my science and the very essence of my humanity too without sacrificing anything about me demonstrating what true authenticity meant.

With an open mind and closer look, one will quickly realize my thesis reads something kind of like a really good book.

However, unlike all the successful experiments captured in this composition, this champion wasn’t built under the most promising conditions.

Against all odds I made all my dreams come true and I managed to accomplish it while fighting innumerable obstacles and all of the stupid racism too.

They told me I can't, I told them I am, and then showed them I CAN.
One cannot go against the grain when it's meant to be; so, challenging me and hating me was ultimately only an exercise in futility.
I am BLACK, BRILLIANT, with a BEAUTIFUL MIND; a social justice warrior supremely crafted through God's intelligent design.
Once upon a time, I was just a kid from the hood talking about a cure for cancer; but now I'm a scholar, discovering the truth, creating knowledge, and presenting to the audiences my most close approximations to the real answers.
Let this be a testament to all those kids like me; make life beautiful, don't ask, seize every single opportunity.
There really isn't a place you can't go; so don't let anyone tell you your destiny is out of your control.
I dream, I believe, I create, and I succeed, so I can surround myself with amazing individuals, greatness, and all of the finer things.
Now I am with all of God's angels, sitting pretty in Washington D.C., but I guess I'm Hollywood because I'm enjoying all of what my hard work brings.
Mama, I wrote all this to say, I made it, these moments they magic, and I just hate that you are no longer around.
The only solace I take is knowing you are up in heaven, watching over me, smiling, and looking down.
Mama, the folks back home say all these degrees done made me change.
Well, what can I say, yeah they are right I did change, what they expect that I should be touting blue around like a badge of honor still saying, "gang gang"?
Mama, I changed for the better despite their belief, and make no mistake I stayed true to my roots in all of my pursuits.
Real talk, I stay sitting in boardrooms, rocking a clean white tee, some retro Jordans, or some fresh Nikes, just like you taught me
So as your heavenly gaze meets my earthly place, please know I've kept every single promise my young heart vowed.
You see I made it up out the streets and It's Dr. Melody Sanders now.
It's safe to say I've earned it the right way, so I hope I've done you proud.

For Beza

(The One That Holds My Life in Her Hands, My Very Best Friend)

To My Very Best Friend Miss Bezawit Addis,

You helped me through the anger, chased away my fears; held me through the sadness and kissed away my tears.

I prayed for you before we even met, not knowing who you'd be; I asked the Lord to send a friend, one chosen just for me.

I asked that they be Godly with the wisdom of his ways; a friend to help and guide me in the troubles of these days.

So when I count my blessings at the end of each and every day, there is only one thing left to say and do; that is to thank God for the very special friend that I have found in you.

There is a very special quality of having a friendship like this that is so pure, deep, and strong; that brings about a special peace of knowing we belong.

You share my dreams, encourage me until my worries lift, and understand me when I cannot find the words to say; so I can't think of a more deserving person to share with all my special days.

You paint my world with joy even when it's filled with broken dreams, you explain it all so clearly when nothing is as it seems.

You are the protector of my heart and the bliss upon my shore, you are a beautiful soul and a life that gives, a life that restores.

You are the one who makes enormous sacrifices, always trying to put me first.

The one who lets me test my broken wings when I insist, no matter how much watching it makes you hurt.

And when your stubborn friend chooses to continue to overextend; you never lose your patience, healing me back to health time and time again.

Only a heart as kind as yours would give so unselfishly.

So, I know God placed a slice of heaven on earth, an angel right in front of me.

There really is no dark place that your perfect love cannot fill.

So, when the world starts causing waves it is your love that makes them still.

You are the mirror of my better self and verifier of the best in me.

Beza, you are my very best friend and the one I love so very deeply.
God made you an amazing person to share gifts with both big and small;
but never forget that you, yourself have been my greatest gift of all.
Love,
Mel

Table of Contents

List of Figures

List of Abbreviations

ABD	Activator Binding Domain
AML	Acute Myeloid Leukemia
BLM	Black Lives Matter
CBP	CREB Binding Protein
CR	Conserved Region
Cryo	Cryogenic
Da	The
DBD	DNA Binding Domain
EM	Electron Microscopy
FBS	Fetal bovine serum
GTF	General Transcription Factor
HDM2	Human Double Minute Two
HIF1α	Hypoxia Inducible Factor 1 Alpha
IDR	Intrinsically Disordered Region
KIX	Kinase Inducible Domain Interaction Domain
NCBD	Nuclear Coactivator Binding Domain
NMR	Nuclear Magnetic Resonance Spectroscopy
NRID	Nuclear Receptor Interaction Domain
MDM2	Mouse Double Minute Two
Myc	Myelocytomatosis
MED19	Mediator Subunit Nineteen
MED28	Mediator Subunit Twenty-Eight
MEF	Mouse Embryonic Fibroblast
PIC	Pre-initiation Complex

PPI	Protein-protein Interaction
RNA	Ribo-nucleic Acid
RN Pol II	RNA Polymerase II
TAD	Transcriptional Activation Domain
TAZ	Transcriptional Adapter Zinc Finger
TCEP	Tris (2-carboxyethyl)phosphine
TF	Transcription factor
2D	Two Dimensional
3D	Three Dimensional

Abstract

The process of transcription is essential to all life, controlling critical physiological processes such as cell growth, maintenance, development, and differentiation. With such a widespread influence on basic cellular processes, regulation of transcription is a high priority, with dysregulation of transcription being a major contributor to several diseases, including cancer, cardiovascular defects, and neurological disorders. While the process of the transcription "reaction" *per se* is quite simple, requiring essentially only an RNA polymerase enzyme, a DNA template and the appropriate nucleotide triphosphates, in cells transcription is much more intricate. Transcription relies on a specialized assembly of a host of dynamic and transient protein -protein interactions (PPIs) acting in a highly coordinated and regulated manner. At the crux of this assembly are transcriptional activators and coactivators. Remarkably, a given transcriptional activator or coactivator will use the same binding surface to complex with a myriad of binding partners including other transcription factors, coactivators, members of the transcriptional machinery and/or chromatin remodeling complexes. Thus, a central question that arises is what are the molecular recognition principles and mechanisms that govern the assembly of a given complex and how is selectivity in binding achieved for a given functional outcome.

A clue to this query may reside in the presence of the intrinsically disordered regions (IDRS) within one or both binding partners present within these kinds of PPIS. For instance, in a given activator, disordered regions within their transcriptional activator domain (TAD) can adopt a distinct tertiary structure to suit the topology of the specific activator binding domain of a coactivator. On the flip side, a given coactivator activator binding domain (ABD) can undergo conformational remodeling fine tuning its docking surface to accommodate different protein binding partners.

This structural adaptability and plasticity enables both specificity and promiscuously whilst at the same time bestowing seemingly unlimited regulatory potential of biological outcomes within the cell. Focusing on the master coactivators, CBP and p300 and their complexes with the prototypical transcriptional activator p53 and using a combination of classical biochemical methodologies and single particle electron microscopy (EM), the goal of this dissertation work has been to test the long-held hypothesis that multi-domain coactivators such as CBP and p300 undergo large conformational changes upon binding to activators.

In chapter two of this dissertation, I describe both the biochemical and biophysical characterization of the intact activator-coactivator complex formed between the coactivators CBP and p300 and the activator p53 and initial single particle EM studies that not only support the feasibility of studying this complex by cryo-EM but also reveal the general architecture of CBP and the accompanying architectural rearrangements that occur upon complex formation with p53. Structural studies on the CBP•$(p53)_4$ complex are expanded upon in chapter three of this dissertation which focuses on the characterization of this complex at higher resolutions by cryo-EM. Taken together, these structural studies have allowed for testing of the outstanding hypothesis that binding at one or more ABDs within CBP or p300 elicits global architectural rearrangements and provides evidence challenging long held beliefs that activators are the predominant flexible partner in transcription complexes.

CHAPTER ONE

Against All Odds, Black in a White House

Introduction

Keep it Moving: Transcription Activation through the Formation of Structurally and Conformationally Dynamic Protein Assemblies.

1.1 Overview of Transcriptional Regulation

The complexity of living organisms is driven by the correct establishment of genetic programs. Notably, control over which genes to express and to what extent dictates core biological processes ranging from organismal development and cell differentiation to tissue homeostasis and the cellular response to stress. Transcription, the critical initial stage in gene expression, can be defined simply as the process of transcribing the information from one strand of DNA into RNA by the enzyme called RNA polymerase II (Poll II). While transcription may be an enzymatic process, its orchestration requires the precise spatial and temporal arrangement of dynamic protein complexes at individual promoters of a given target gene. These complexes consists of chromatin modifying and remodeling enzymes and the general transcription factors (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH) that work in concert to relax DNA up and downstream of the gene promoter sequence, direct RNA polymerase II to the transcription start site and form the preinitiation complex (PIC)[1, 2]. While all of these aforementioned proteins represent core PIC components that are required for the transcription of nearly all genes, transcriptional activation cannot proceed without two distinct classes of protein termed transcriptional activators and co-activators **(Figure 1.1).**

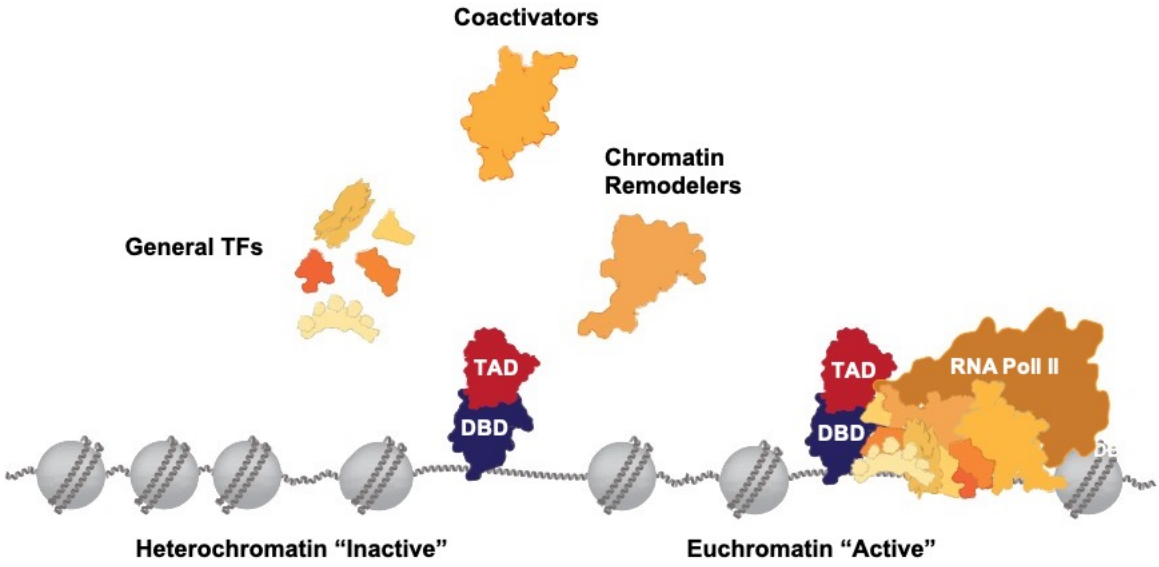


Figure 1.1 General Model of Transcriptional Activation. In transcription, gene specific activators bind DNA through their DNA Binding Domain (DBD) and recruit other regulatory proteins through their transcriptional activation domain (TAD). The other proteins include activators, general transcription factors, and chromatin remodelers who all work together to initiate the process of transcription.

Therefore, it is the cooperative engagement of an elaborate network of protein-protein interactions (PPIs) between the general transcription machinery, activators, and coactivators that work in an interconnected fashion to regulate correct gene expression patterns. Due to the role that these many PPIs play in maintaining a healthy cellular environment, it therefore should come as no surprise that their dysregulation underpins disease. Thus, there is a strong need for a molecular-level understanding of the assembly process to facilitate drug discovery.

1.2 Eukaryotic Transcriptional Activation Is Achieved through PPIs between Transcription Activators and Coactivators

Transcription Activators "Activate" Transcription

The process of eukaryotic transcription is a complex task. This task is achieved through an intricate network of proteins that work together to orchestrate the correct establishment and propagation of genetic programs. Several lines of evidence have demonstrated that for nearly all eukaryotic transcriptional programs, recruitment of the transcriptional machinery to specific genomic loci is mediated by a distinct class of

proteins known as transcriptional activators [2-5]. These proteins recognize and bind to specific sequences in the promoters of cognate genes and recruit a suite of coregulatory proteins that modulate the transcriptional output of these genes by modifying chromatin structure and assembling the RNA Polymerase II (RNAPII) pre-initiation complex. To carry out this role, transcription activators function through modular domain architectures, composed minimally of a DNA binding domain and transcriptional activation domain (**Figure 1.2).** Accordingly, the DNA binding domain of the activator provides the specificity for its action (in terms of which gene to activate), whereas its activation domain is responsible for forming dynamic, transient PPIs to assemble the machinery.

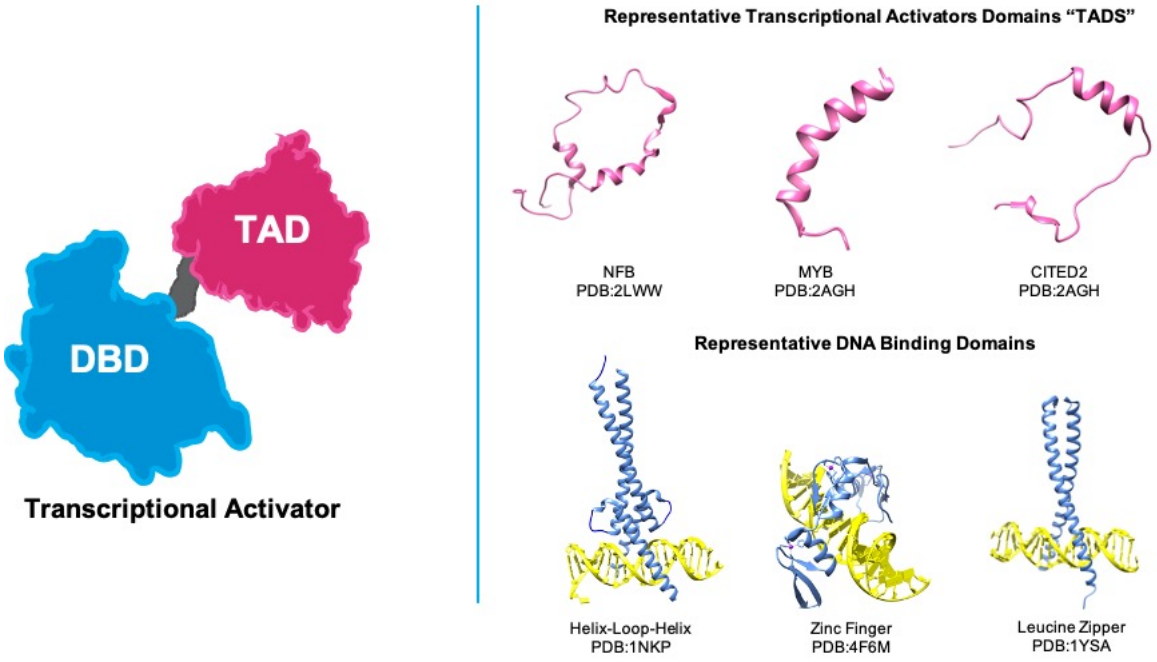


Figure 1.2. Schematic Showing the Modular Nature of Transcriptional Activators. A schematic representation of the modular nature of transcriptional activators is presented in the left panel showing the DNA bind domain in blue and the transcriptional activation domain in magenta. Several structures of representative TADs (pink) and DBDs (blue) are shown on the right.

The DNA Binding Domain (DBD)

Activators utilize DNA binding domains (DBDs) to activate gene expression through sequence specific DNA binding in target genes upstream of the transcription start site. DBDs can be grouped into different structural motifs, based on the way they recognize double-stranded DNA. Structural motifs of DBDs identified to date include zinc fingers, helix-turn-helix, leucine zipper, and helix-loop-helix motifs [6]. Although less common, two smaller DNA binding domain families have been identified that use beta-

sheets for binding. In spite of the apparent structural diversity of the DNA binding domains some general patterns and principles have emerged for site-specific recognition. For instance, the aforementioned DNA binding domain motifs and many others recognize short, specific DNA sequences through elaborate contacts made between alpha-helices in the DNA-binding domain and bases in the major groove of the DNA double helix [7]. The secondary structures of the DBD are orientated in such a way that allows for intermolecular contacts between the critical side chains of amino acids, and specific bases of DNA and the phosphate backbone with additional specificity for a particular promoter sequence provided by distinct patterns of hydrogen bonding [7]. Consequently, transcription activators achieve in many cases an exquisite level of specificity for DNA recognition through subtle differences in the spatial orientation of secondary structures positioned within the DNA major groove, the physiochemical characteristics of interaction surfaces, and the specific patterning of hydrogen bond networks.

The Transcriptional Activation Domain

In addition to a DNA binding domain (DBD), a typical activator also contains a transcriptional activation domain (TAD) that is required for the activator to stimulate transcription. Following correct genomic localization promoted by DBD of activators, the transcriptional activator domain (TAD) recruits coregulatory protein complexes or the transcriptional machinery through a network of protein-protein interactions [3, 5, 8-10]. Unlike DNA binding domains that require elaborate structures for DNA recognition, minimal activation domains tend to be short protein sequences often with very limited sequence complexity. Owing to their lack of sequence similarity and their low-complexity sequences, historically these domains have been classified according to their amino acid profile as acidic, glutamine-rich, proline-rich or serine/threonine-rich [3-5, 8-10]. Furthermore, in isolation TADs tend to be intrinsically disordered, often existing as extended random coils or collapsed globules, only adopting α-helical secondary structures upon binding to cellular targets [10-13]. The intrinsic lack of order within TADs is hypothesized to provide functional advantages including (1) a degree of flexibility, to interact with different proteins on different occasions, (2) accessible sites for post translation modification allowing for strategic, tunable, and reversible regulation of their function, and (3) the ability to bind proteins with high specificity but modest affinity [11,

13-15]. In other words, the intrinsically disordered regions (IDRS) within TADS provide specific interactions while at the same time allowing rapid and spontaneous disassociation and termination of a signal when the need arises. Taken together, these tunable structural characteristics allow activators to interact with multiple components of the transcriptional machinery, as well as other proteins, allowing the same polypeptide to undertake different interactions with distinct transcriptional outcomes.

Coactivators are Transcriptional "hub" Proteins in da Hood

Despite important roles that transcription factors play in transcriptional activation, they are not the only actor involved in the regulation of this important cellular function. Another important group of proteins required for transcription are coactivators. Coactivators are often referred to as transcriptional hub proteins, both because they reside at the center of transcriptional regulatory networks and they function through binding to many partners. These binding partners can include RNA polymerase, other coactivator and coactivators complexes, as well as enzymes and enzyme complexes that carry out the deposition and removal of post-translational modification in addition to a multitude of transcription factors (activators)[16]. The traditional model of coactivator function is that they simply serve as somewhat static molecular bridges connecting DNA-bound transcriptional activators and the remainder of the transcriptional machinery. However, emerging structural and mechanistic evidence indicates a more complex model in which binding-induced conformational changes of a coactivator or coactivator complex allosterically regulates key aspects of transcription initiation. Early structural studies of the *S. cerevisiae* Mediator coactivator complex, for example, showed a significant architectural change upon binding to transcriptional activators that altered the binding surface for RNA polymerase II [2, 17, 18].

As described earlier in eukaryotes, RNA polymerase II (pol II) transcribes all protein-coding genes in addition to many noncoding RNAs. Whereas there are many determinants that contribute to the regulation of pol II activity, the Mediator complex represents a critical transcription coactivator hub protein that is required for expression of most, if not all, pol II transcripts. In metazoan cells, multiple pathways that are responsible for homeostasis, cell growth, and differentiation converge on Mediator through transcriptional activators and repressors that target one or more of the almost 30 subunits

of this complex [2, 17-21]. Specific interactions occur both between individual Mediator subunits and transcriptional activators and between Mediator and Pol II. Besides interacting directly with RNA polymerase II and facilitating activator function, recent findings show that Mediator influences nearly all stages of transcription including initiation, elongation re-initiation, transcriptional memory, and coordinates these events with concomitant changes in chromatin organization **(Figure1.3)**. The functional versatility and multifaceted role of Mediator in gene expression further bolsters the idea of Mediator as a transcriptional hot hub.

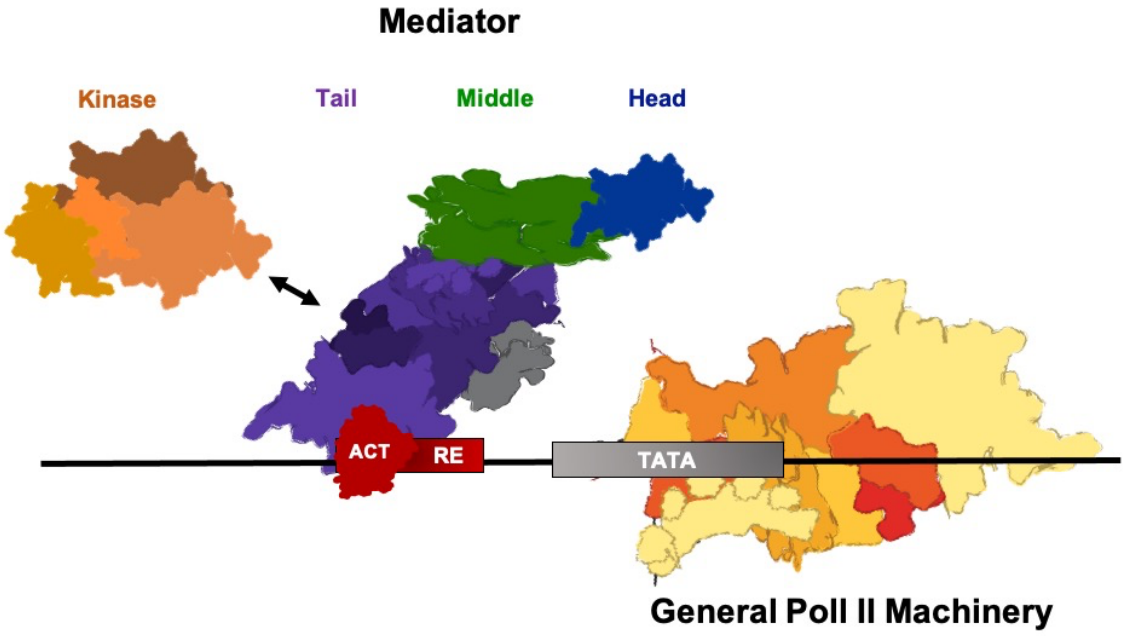


Figure 1.3. A Simplified Model of The Mediator Complex and its Role in Transcription. Shown are Mediator modules head, middle, tail, and the exchangeable kinase module are colored in blue, green, purple, and orange respectively. Mediator serves as a central regulator invoking a 'molecular bridge' between gene specific activators (ACT) bound to regulatory DNA elements (RE) that allows it to integrate and communicate regulatory signals from DNA-binding TFs directly to the Pol II enzyme and the rest of the PIC

The A-B-Ds (Activator Binding Domains) of the Activator Alphabet

Activator interactions with coactivators such as Mediator occur via activator binding domains (ABDs) which have now been shown to be conformationally dynamic protein folds that exist within coactivator proteins and enable interaction with the multitude of transcriptional activators [22-25]. The activator binding motifs of transcriptional coactivators are structurally diverse and conformationally dynamic. In contrast to the

transcription activator domains most activator binding domains within transcriptional coactivators exist as more structured motifs. **(Figure 1.4).**

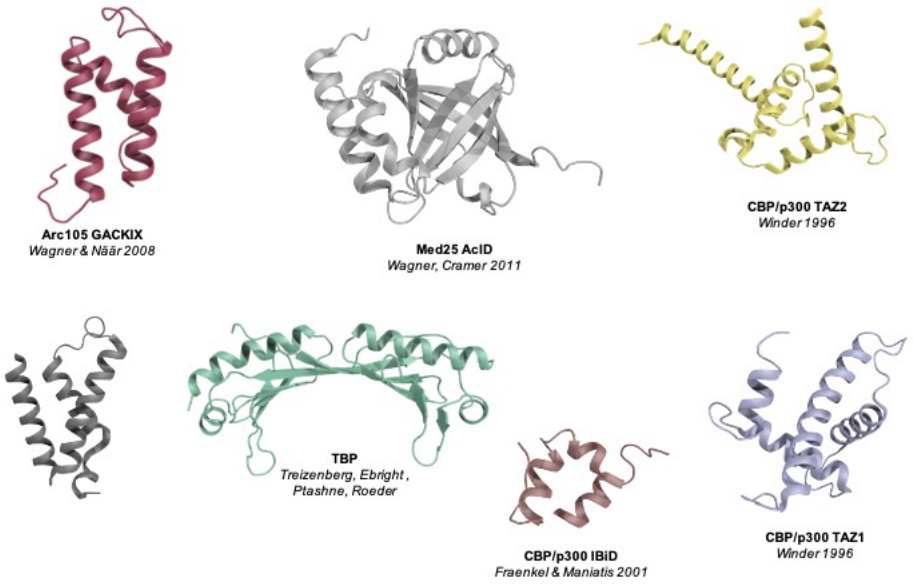


Figure 1.4. Diversity of Activator Binding Domains (ABDS). The majority of ABDs studied to date possess high degree of alpha-helical structure and are structurally diverse adopting a number of unique conformations.

As described later KIX, represents one of several conformationally plastic domains found in the master coactivators CBP and p300. KIX interacts with more than 15 transcription activators at two distinct binding interfaces [21]. Conformational changes within KIX upon partner binding at one site can significantly alter binding partner preferences at the second site leading potentially to tailored downstream effects in the transcriptional outcome [26]. Thus intriguingly, KIX functions not only as a structured scaffold for binding of disordered ligands but also as an allosteric regulator of transcription **(Figure 1.5).** Another excellent example highlighting not only the structural diversity, but also the dynamic nature characteristic of ABDS is the Acid domain of Med25, a member of the Mediator complex. Med25 contains an ABD fold unique to all other known coactivators bearing helices and loops that display significant conformational mobility and a seven-stranded β-barrel core that is not structurally flexible.

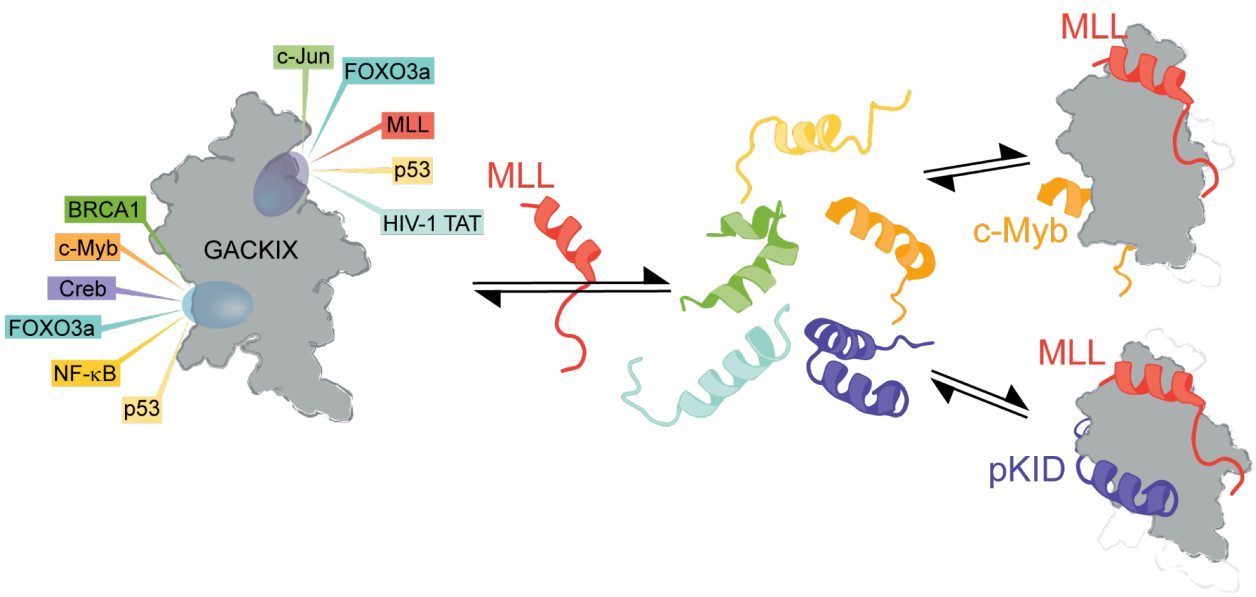


Figure 1.5. Conformational Plasticity in KIX. The KIX domains of CBP and p300 are able to bind more than ten activator proteins due to their conformational plasticity,

1.3 Dysregulation of Transcriptional Function Contributes to Diseases

Transcription of protein-coding genes in eukaryotic cells is regulated by an ensemble of proteins, whose central component is the enzyme pol II. To accomplish its action efficiently, pol II requires a battery of accessory proteins including transcription factors (activators), coactivators, and chromatin regulators. The critical importance of these proteins for cell physiology is illustrated by the observation that mutations or aberrations in their biochemical properties or the regulatory mechanisms that fine-tune their activity, are frequently associated with various human diseases [8, 10, 12, 14, 16, 27-29]. For instance, human malignancies ranging from cancer to diabetes, to various neurological and developmental disorders, to infertility and obesity have all been shown to be a direct consequence or associated with transcriptional misregulation. For example, the Myc protein is a transcription factor that regulates a variety of cellular processes including cell growth and proliferation, cell-cycle progression, transcription, differentiation, apoptosis, and cell motility. Deregulated expression of c-MYC occurs in the majority of human cancers where it plays a pivotal role as a regulator of tumor genesis and cancer progression [30, 31]. Most tumor cells depend on the transcription factor c-Myc for their growth and proliferation**.** *Myc* is the most frequently amplified oncogene, and the elevated

expression of its gene product is associated with tumor progression and poor clinical outcome.

Mutations in the Mediator coactivator complex have also recently been implicated in the development of various tumors. As described earlier, Mediator is an essential coactivator complex that acts as a bridge between transcription factors bound at upstream regulatory elements and the transcription machinery. In this capacity, Mediator serves to channel regulatory signals from activator and repressor proteins to affect changes in gene expression programs that control diverse physiological processes, including cell growth and homeostasis, development, and differentiation [17, 19, 20]. Importantly, the pre-initiation complex (PIC) consists of Mediator, Pol II and TFs, with the mediator complex serving as a central scaffold within the PIC and a regulator of Pol II activity. Due to the importance of mediator's role in the transcription of eukaryotic genes, it is conceivable that disruption of its correct function is expected to have huge pathogenic consequences. Unsurprisingly, in the last years a number of subunits, other than the complex itself, have been suggested to have a role also in tumorigenesis [19, 20, 32, 33]. Indeed, immunohistochemical studies have shown the expression of distinct subunits in several solid tumors. For instance, MED19 has been found up-regulated in human bladder cancers compared with adjacent benign tissues [19]. Similarly, a high expression of this subunit has also been shown in other tumors, like lung tumors, hepatocellular carcinoma, and osteosarcoma [19]. Another example shows that increased expression of MED28 has been found in breast cancer.

In summary the Mediator complex and the aforementioned c-MYC illustrate how proteins and their resulting complexes can serve as convergence points in oncogenic signaling pathways and how they become functionally altered in many cancers. This makes them attractive targets for anticancer therapeutic agents. However, an ongoing challenge that remains is a lack of in-depth understandings of the relevant protein-protein interaction networks that are misregulated in these diseases. This can only be achieved through a structural, functional, and mechanistic understanding of these molecular players. It such an understanding that will aid in the development of high-quality therapeutics.

1.4 The Master Coactivators CBP and p300 are Paradigms for Understanding Transcriptional Regulation

The Center of Attention: CBP and p300 Serve as Central Nodes in Eukaryotic Transcriptional Regulatory Networks

CBP and p300 are master transcriptional coactivators that integrate numerous signaling pathways and play critical roles in cell proliferation, differentiation, apoptosis, and DNA repair [16, 22, 27]. The transcriptional regulatory properties of CBP and p300 appear to be exerted through two main mechanisms. The first mechanism is by bridging transcription factors and the general transcription machinery through PPIs [15]. The second mechanism by which CBP and p300 facilitate transcription is by acetylating histones in the vicinity of target promoters, forcing chromatin into a more open and accessible configuration **(Figure 1.6**) [15, 16, 27, 34, 35].

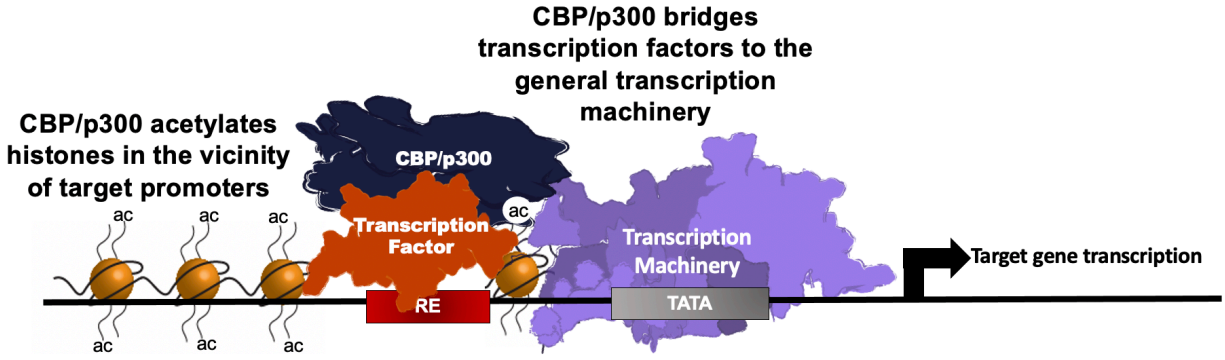


Figure 1.6. Mechanism of Transcriptional Activation by CBP/p300. CBP and p300 deposit acetylation marks near target promoters remodeling chromatin to be in a more active state. As coactivators they bind transcription factors and other members of the transcriptional machinery to help stimulate transcription.

Structurally, CBP and p300 are very large (~265 kDa) modular proteins sharing several conserved folded domains which include the nuclear receptor interaction domain (NRID), a KIX domain, three cysteine-histidine rich domains (TAZ1/CH1, CH2, TAZ2/CH3) and the nuclear coactivator binding domain (NCBD) also referred to as the IRF-3 binding domain (IBID)[15, 16, 22, 36]. Additionally, CBP and p300 each contain a histone acetyltransferase (HAT) domain, endowing them both with intrinsic acetyltransferase activity, and a bromodomain that binds acetylated lysines as well as a PHD finger motif with unknown function **(Figure 1.7)**. Large regions predicted to be

intrinsically disordered connect the more structured domains and account for approximately sixty percent of the entire protein. The TAZ1, TAZ2, KIX, and NCBD domains are conformationally dynamic ABDS and form the interaction sites for more than 400 eukaryotic transcription factors and other regulatory proteins and are also often the target of viral oncoproteins [15].

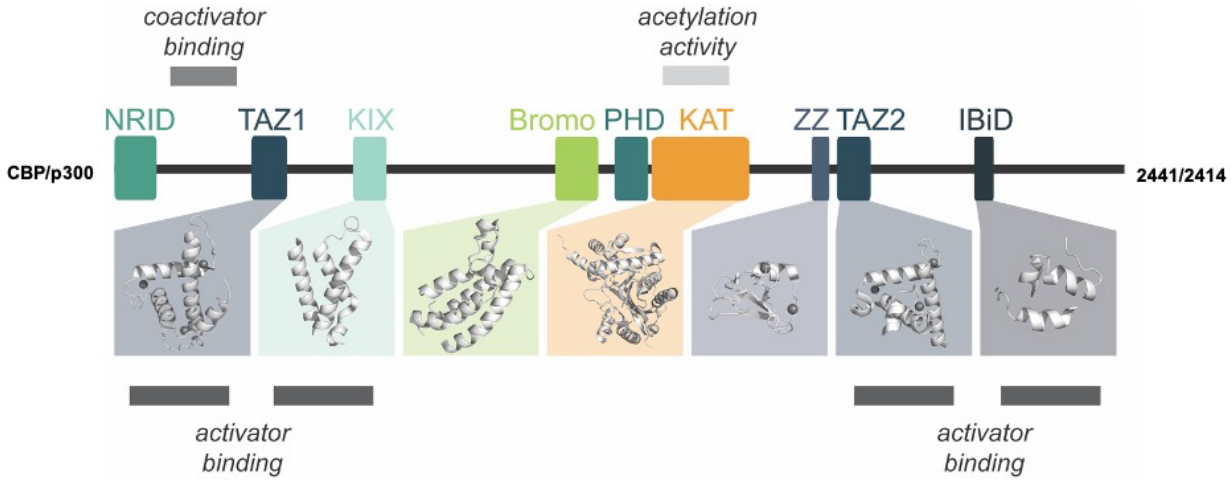


Figure 1.7. Domain Architecture of CBP and p300. Linear representation of the full-length CBP/p300 proteins showing regions and domains that are highly conserved. Shown are the NRID, nuclear receptor interaction domain; TAZ1 and TAZ2, transcriptional adapter zinc binding motifs; KIX; Bromo, bromodomain; CH2, cysteine-histidine-rich domain 2; KAT, lysine acetyltransferase domain (or HAT histone acetyltransferase domain) including a disordered regulatory loop; ZZ, dystrophin-like small zinc binding domain; and NCBD(IBiD), nuclear receptor co-activator binding domain.

Taken together, the presence of more structured domains with intervening disordered regions is proposed to afford CBP/p300 plasticity for accommodating the binding of a large repertoire of partners, making them ideally suited to function as molecular hubs in protein-protein interaction networks. For instance, CBP and p300 have been shown to associate with the basal transcription machinery and other coactivators in addition to the large gamut of activators mentioned above [16]. An on-going conundrum in the field has been to understand how activator binding domains such as those present in CBP and p300 recognize with precise specificity such a broad range of activators with disparate sequences to control spatially and temporally regulated transcription programs. Focusing on the activator binding domains of CBP and p300 below I will review some generalities of the structural principles of CBP and p300 that guide transcription factor interactions.

1.5 Know Your Rights and Your ABDs: The Structural Basis of CBP and p300 Transcription Factor Interactions

KIX "pKID tested MLL approved"

The KIX (kinase-inducible) interaction domain was discovered by Parker et al. in 1996 in mouse CBP, as the specific and minimal region that was sufficient to bind and interact with phosphorylated CREB and then activate transcription [21]. KIX comprises a highly conserved, independently folding three-helix bundle containing two alpha helices and one 3_{10} helix that serves as a primary docking site for transcription factors that function in hematopoietic differentiation including CREB, C-Myb, MLL (mixed lineage leukemia protein), c-Jun, E2A, and FOX03, among others [21, 37]. KIX consists of two well-defined binding sites often referred to as the "MLL site" and "pKID/c-Myb site" named after the well-known activators that bind to these respective locations.

Since its initial identification, a number of solution structures of KIX alone or TAD-KIX complexes have been determined revealing a surprising degree of diversity in binding poses [21, 37-40]. The intrinsically disordered activation domains of CREB and c-Myb fold into helical conformations and bind in a shallow hydrophobic groove in the α1-α3 helix surface of KIX [15]. The kinase-inducible activation domain of CREB (KID), phosphorylated at Ser^{133} (pKID), forms a pair of orthogonal helices upon binding to KIX, with the C-terminal helix dominating the binding interaction. c-Myb on the other hand forms a slightly bent helix that binds in the same site as the pKID helix [15]. In contrast, the MLL activation domain binds in a helical conformation on the opposite face of KIX, in a hydrophobic groove formed by the α_2, α3, and 3_{10} helices. MLL and Myb, or MLL and pKID, can bind simultaneously and synergistically to KIX to form a ternary complex, resulting in allosteric enhancement of binding affinity [15]. Thus, binding of one TAD to KIX significantly influences the binding partners that can bind at the other site and ultimately influencing the interactome of CBP and p300. Furthermore, binding to each site remodels the KIX structure in a number of ways enabling distinct association modes, further highlighting the structural plasticity of KIX **(Figure 1.8)**.

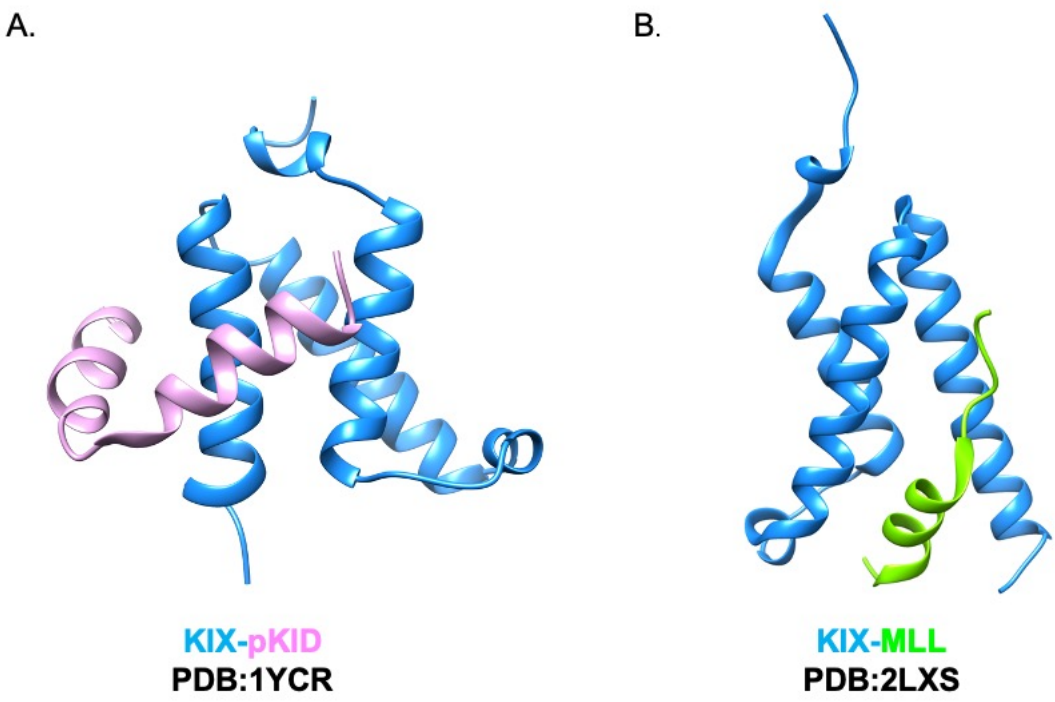


Figure 1.8. **Conformational Remodeling of the KIX Domain.** Conformational remodeling of KIX domain results in significant differences in overall 3D arrangements as highlighted by **(A)** the KIX-pKID complex (left panel; blue, pink) and **(B)** the KIX-MLL complex (right panel; blue, green).

Interestingly, KIX can also bind activation domains with bipartite interaction motifs. For example, FOX03, a transcription factor that regulates cell differentiation, survival, and apoptosis genes, contains two amphipathic interaction motifs that bind simultaneously to both KIX sites [16]. The two amphipathic TADS termed CR2 and CR3 of FOX03 can associate with KIX in two different modes: CR2 bound to the MLL site and CR3 bound to the c-Myb site and conversely CR3 bound to the MLL site and CR2 bound to the c-Myb site [16]. The binding of FOX03 to KIX further demonstrates how malleable this domain is.

TAZ1 and TAZ2 domains

The TAZ1 and TAZ2 (transcriptional adaptor zinc finger) domains share a similar fold that consists of four alpha helices stabilized by three zinc atoms. The largest structural differences between these two motifs are the length of α-helix$_1$ which is extended in TAZ1 and the orientation of α-helix$_4$ which is rotated 180° between the two domains. In spite of their topological likeness, TAZ1 and TAZ2 differ extensively in their

amino acid sequence and are highly selective in their interactions binding different subsets of intrinsically disordered activation domains.

The TAZ1 motif facilitates the transcription of genes involved in the immune and inflammatory response, cellular proliferation, survival, and the hypoxic response. IDRS that bind this motif with strong affinity tend to be relatively long, making extensive hydrophobic contacts in a deep binding groove through the presence of their multiple amphipathic regions [15, 25, 41-43]. (**Figure 1.9 A-B)**. These contacts are further supplemented by electrostatic interactions between acidic residues in the ligand and positively charged side chains on the surface of TAZ1.

The binding of HIF-1α to TAZ1 represents one very nice example of a TAD-TAZ1 complex. Transcriptional regulation by HIF-1α is entirely dependent on the interaction between the Hif-1C-terminal activation domain (CAD) and the TAZ1 domain of CBP/p300. The TAZ1 domain recognizes the CAD of HIF-1α, which functions to restore cellular oxygen homeostasis by inducing transcription programs for crucial adaptive genes. Under normoxic conditions, hydroxylation of an asparagine residue occurs and inhibits binding to CBP or p300[15, 25, 41-43]. In hypoxia, asparagine hydroxylation is blocked and recruitment of the CBP or p300 coactivators occurs. In isolation, the CAD is intrinsically disordered and remains relatively extended upon binding, encircling almost entirely around the TAZ1 domain [41, 43, 44]. Three short helices are formed upon binding and are stabilized by a network of intermolecular interactions. The Asn-803 side chain, which functions as the hypoxic switch described above, is located on the second of these helices and is buried in the molecular interface [41, 43, 44]. The third helix of the HIF-1α CAD docks in a deep hydrophobic groove in TAZ1, further providing intermolecular hydrophobic interactions that contribute to the stability of the complex (**Figure 1.9 A)**. The affinity of the HIF-1α-TAZ1 complex is unusually high with a dissociation constant of 7 nM.

Unlike TAZ1, TAD interactions are mostly localized to the surface of TAZ2 at a hydrophobic helical interface consisting of α1, α2, α3, and α4 helices [24, 45-48]. The TAD of STAT1 binding to TAZ2 and that of E1A provide excellent illustrations of this binding mode **(Figure 1.9 C- D)**. Furthermore, the recruitment of STAT proteins by both TAZ1 and TAZ2 provides a superb example of selectivity of the TAZ1 and TAZ2 domains.

CBP TAZ1 and TAZ2 specifically recognize the TADS of STAT2 and STAT1, respectively with over 100-fold selectivity.

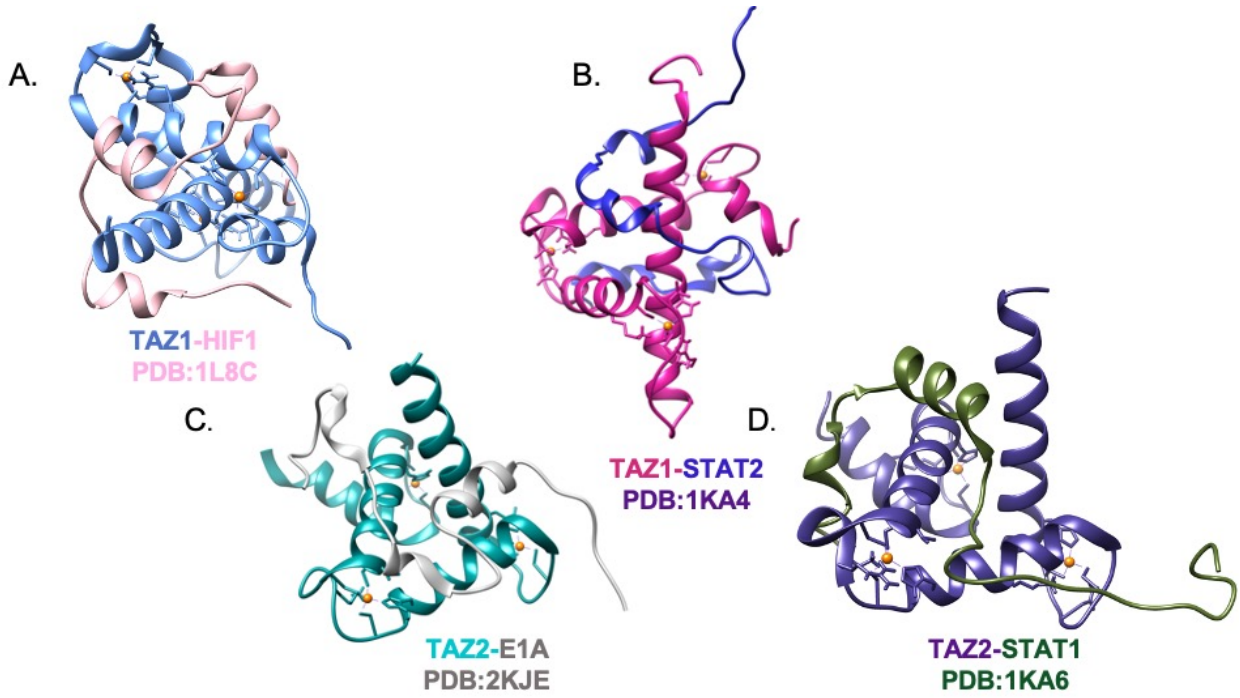


Figure 1.9. Structures of TAZ1 and TAZ2 complexes. TAZ1 and TAZ2 domains share a similar fold that consists of four alpha helices stabilized by three zinc atoms. **(A-B)** TAZ1 primarily binds long, extended activation domains in deep grooves whereas **(C-D)** TAZ2 binds tightly to discrete amphipathic helices through an exposed hydrophobic surface.

"Da" NCBD

In the unbound state the nuclear coactivator binding domain of CBP and p300 (also known as IBID) can be described as a molten globule presenting transient helical structure [49]. Intriguingly, interactions between NCBD and the unstructured TADS of other activators involve the synergistic folding of two intrinsically disordered polypeptides. Upon association with its physiological binding partners, the NCBD folds to form a bundle of three helices that differ significantly in their topological arrangements in complexes [49-51]. For example, NCBD folds into two remarkably different states depending on the ligand being ACTR or IRF-3 **(Figure 1.10)**. The complexes formed by the NCBD domain of CBP and p300 provide an exemplary demonstration of ABD conformational remodeling and conformational plasticity.

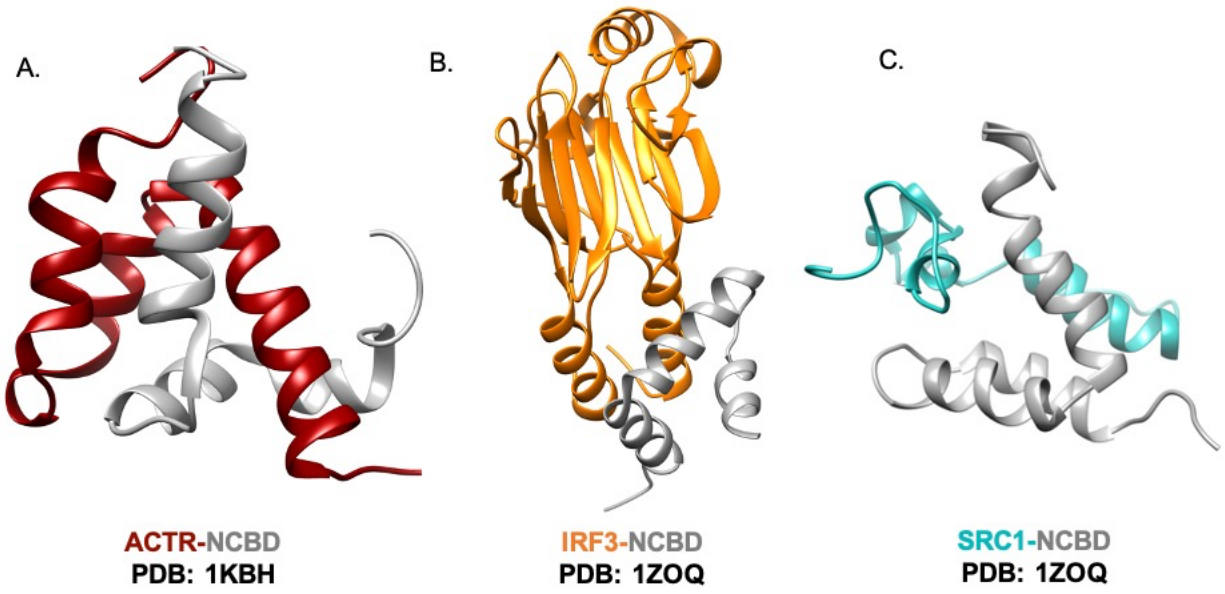


Figure 1.10. Various TAD-NCBD Complexes. NCBD is a paradigm for conformation plasticity. As shown complex with **(A)** ACTR **(B)** IRF3 and **(C)** SRC1, NCBD adopts unique conformations in all.

Allies "Activating" Activism: the CBP and p300 Catalytic Core

CBP and its closely related paralog p300 are coactivators and lysine acetyltransferases that interact with and acetylate transcriptional regulators and the chromatin landscape leading to changes in chromatin structure. CBP and p300 are both modular in their architectures and contain several well-defined domains including TAZ1, KIX, TAZ2, and NCBD as noted above. CBP and p300 also contain a HAT (histone acetyltransferase) domain which has been shown to acetylate multiple lysines on various histones [36, 52-55]. Acetyltransferases (HATs) such as the ones found in CBP and p300 transfer an acetyl group from the metabolite acetyl-CoA to the ε-amino group of lysine residues. Lysine acetylation provides binding sites for effector proteins containing bromodomains and histone deacetylases (HDACS) remove such chromatin decorations, creating a discrete series of actions including the writing, reading, and erasing of histone acetylation. This allows for a mechanism for differential regulation of gene transcription.

With regard to CBP and p300, they been shown to acetylate K14, K18, K23, and K27 on histone H3, with H3K27 acetylation being a well-studied mark correlated with transcriptionally active states of chromatin [52]. On histone H4, CBP and p300 acetylate K5, K8, and K12. CBP and p300 also acetylate transcription factors and in doing so

modulate their activity [52]. Notably, it has been demonstrated that the HAT domain of p300 and CBP contain a lysine rich autoregulatory loop (AL) that when hypoacetylated, inhibits acetyltransferase activity of CBP and p300 by competing with positively charged substrates for binding to the active site [53, 54]. Hyperacetylation of the AL displaces this loop and greatly enhances CBP and p300 HAT activity indicating that the AL plays a regulatory role in the enzymatic activity of these protein players.

The HAT domain of CBP and p300 is proximal to other domains including a bromodomain, and the cysteine/histidine regions which have both been proposed to play a role in regulating not only the catalytic activity of CBP and p300 but also their substrate specificity [35]. In common with other bromodomains, the CBP and p300 bromodomains recognize and bind acetylated lysines on histone tails and in non-histone substrates including acetylated lysines in histones H2A, H2B, H3, and H4. To gain further insights into how the bromodomain, CH2 domains, and HAT domains work together to ensure proper function of CBP and p300 and its catalytic activity several groups have determined the crystal structure of 1) the catalytic core of human p300 consisting of the bromodomain, CH2 region (PHD and RING) and HAT domain and 2) the CBP catalytic core comprising the bromodomain, CH2 (PHD and RING), HAT, and ZZ [35, 55].

The structure of human p300 catalytic motif shows that the bromodomain, PHD, RING, and HAT domains form a compact module in which the RING domain is in close proximity to the substrate binding site of the HAT (**Figure 1.11B)** [55]. The electron density observed for the CH2 region (PHD and RING) domain revealed a unique motif. The PHD domain is rigidly attached to the C-terminal subdomain of the HAT motif and interacts with the bromodomain by a network of hydrogen bonds and hydrophobic interactions. The common aromatic cage of prototypical PHD domains used to “read out” the methylation status of H3K4 is absent in the p300 PHD motif. An additional surprising detail of this motif is that the PHD is discontinuous being interrupted by a RING domain. The RING and HAT domains engage through linker L1 and L2 that also join the RING and PHD domains. These linkers converge toward the PHD domain and interact with the HAT principally through electrostatic interactions. One striking feature of the p300 HAT domain is the presence of an uncommonly long substrate binding loop positioned in a manner to influence substrate binding, this loop comes into direct contact with the RING

domain suggesting that the RING motif may play a role in substrate binding and/or catalysis. Indeed, mutational analyses indicate that RING domain deletion leads to a marked increase in HAT activity.

The crystal structure of the CBP catalytic core at 2.4 Å encompassing the bromodomain, CH2 (PHD and RING), HAT, and ZZ domains consists of a central modular core similar to that of p300 in which the bromodomain, PhD, and HAT domains are intimately packed through both electrostatic and hydrophobic interactions **(Figure 1.11A)** [35]. In contrast to the solved p300 HAT structures which were all solved in the presence of acyl-CoA variants, the apo-CBP HAT shows that the acetyl-CoA binding site is fully preformed in the absence of bound cofactors. Moreover, the majority of the active site residues, including the catalytic tyrosine (Y1504 in CBP, Y1467 in p300) adopt nearly identical conformations in both proteins. Some notable primary differences between the structures of CBP and the p300 catalytic cores include the presences of the ZZ domain in the CBP core contact and its omission in the p300 core construct, the weak electron density for the CBP ring domain (likely due to orientational flexibility) and finally the rotation of the bromodomain in CBP. Further inspection of the bromodomain in CBP presents the bromodomain 4 angstroms closer to the HAT domain. Given that the bromodomains of CBP and p300 are almost identical, this orientational difference may account for the difference in substrate selectivity previously observed. Considering that many mutations cluster in the catalytic core of both CBP and p300 structural studies of these regions serve as solid beginnings to understanding not only how chromatin substrate targeting and HAT regulation happens but how dysregulation of HAT activity occurs in disease.

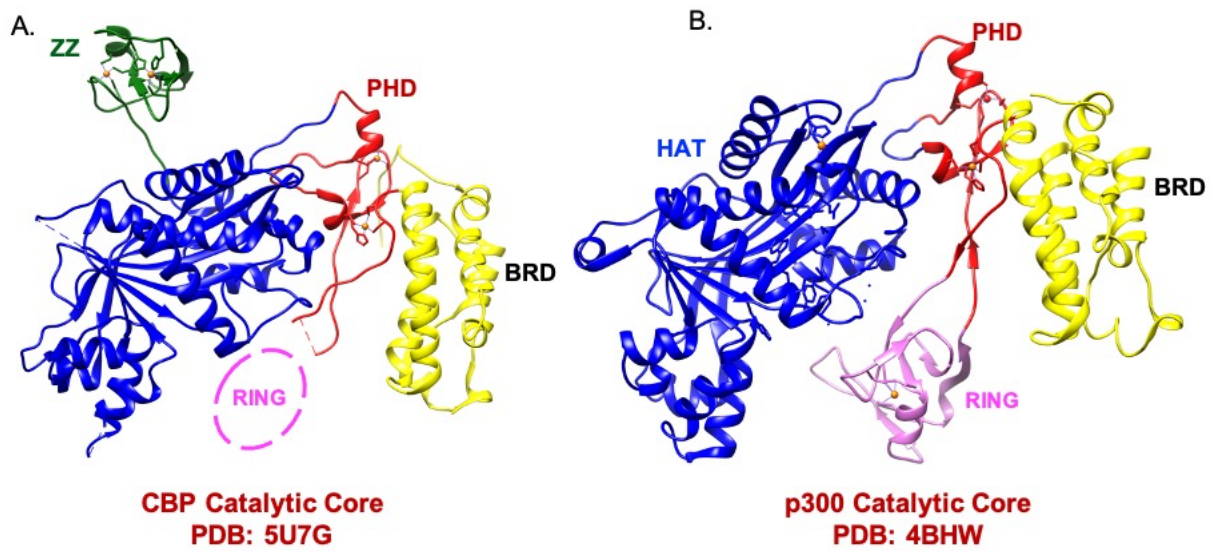


Figure 1.11 The CBP and p300 Catalytic Cores. The structures of the BRD, PHD, and HAT domains of **(A)** CBP and **(B)** p300 are very similar. The primary differences between the structures of the CBP and p300 catalytic cores are the presence of the ZZ domain in the CBP construct and the weak and diffuse nature of the electron density for the CBP RING domain.

1.6 We are a Community in Pain: Dysregulation of CBP or p300 function is associated with Human Disease and Disorders

CBP and p300 play pivotal roles in various signaling pathways and regulate a multitude of cellular processes thus mutations and dysregulation of these proteins have been implicated in a panoply of human diseases including cancer. The critical importance of these proteins is underscored by the fact that CBP homozygous knockout, p300 homozygous knockout, and the p300 and CBP heterozygous knockouts are embryotic lethal [16, 27, 34]. More support for the role of CBP and p300 in maintaining healthy human growth and development come from in vivo studies that demonstrate KIX or CH1 deletion leads to pronounced significant defects in hematopoiesis. On top of that, single heterozygous loss-of function mutations in p300 or CBP, are one of the known causes of Rubinstein Taybi Syndrome (RTS), a congenital development disorder that presents clinically as postnatal growth deficiency, mental retardation, skeletal and cardiac abnormalities, among other ailments [16, 27, 34]. Additionally, another phenotype of RTS is an increased likelihood to develop cancer.

Apart from RTS-associated tumors, loss of function and gain of function in CBP and p300 also contributes to human cancers [16, 27, 34]. CBP and p300 may exert tumor

suppressor effects by acting in concert with of other bona fide tumor suppressors such as the retinoblastoma protein Rb, BRCA1, and FOX03. As coactivators of C-Myb, c-Myc, and AR, p300 and CBP also promote cell proliferation and tumorigenesis in certain contexts. C-Myc in particular is an oncoprotein is highly expressed in cancers of myeloid, lymphoid, and erythroid lineages and studies have suggested that transcriptional activation of c-Myc by CBP and p300 allows this oncogenic transcription factor to sustain the proliferative state of hematopoietic precursors cells. In the case of c-Myb, studies in mice have shown that the c-MYB-CBP or c-MYB-p300 interaction is requisite for leukemic transformation when certain MLL translocations are present [16, 27, 38, 39, 56, 57]. For both c-Myc and c-Myb, CBP and p300 act as co-conspirators colluding with these oncoproteins by aiding in their coactivator driven transcription.

There are also a number of examples of chromosomal translocations associated with leukemia that involve CBP or p300. In these instances, particularly in certain hematological malignancies, part of CBP or p300 is translocated such that it becomes fused with another protein, eliciting inappropriate activity. One example of this occurrence is the fusion protein formed between the mixed lineage leukemia gene with p300 or CBP. This fusion is found with patients with infantile acute leukemia and patients with treatment-related leukemia. Another example fuses the monocytic leukemia zinc-finger (MOZ) gene with CBP and is found in acute myeloid leukemia (AML)[16, 34]. The MOZ drives leukemogenesis through a number of proposed mechanisms. In one scenario, MOZ-CBP has been shown to enhance transcription from NF-κB promoters. MOZ-CBP also suppresses p53 transcription programs, providing yet another avenue for leukemogenesis to occur. Taken together, a there are a number of CBP and p300 mechanisms that can drive human malignancies. However, the precise molecular mechanism that underlies dysregulation and pathogenesis of CBP or p300 appears to differ depending on which domain is mutated. That being said, that are multiple avenues to pursue for the successful therapeutic targeting of CBP and p300 when their function goes awry.

1.7 The tumor suppressor p53 requires coordinate engagement with CBP or p300 for function

p53: The Quintessential Tumor Suppressor

The p53 protein is a potent tumor suppressor that has been intensively studied since its discovery in 1979. Compelling evidence in support of p53's role in tumor suppression was initially provided by findings that show *p53* is mutated in at least half of all human cancers. Further support for the pivotal function of p53 in tumor suppression came from studies of individuals with Li-Fraumeni syndrome, who commonly inherit a mutant *p53* gene and are highly predisposed to cancer development and a characteristic spectrum of tumors including sarcomas, brain cancers, breast cancers, and adrenocortical carcinomas [58-61]. Finally, unequivocal confirmation of the crucial role for p53 in tumor suppression was demonstrated by the completely penetrant cancer phenotype of *p53*- null mice.

P53 restricts tumor development by serving as a sensor of cellular stress and responding to diverse signals, including DNA damage, hypoxia, oncogene expression, nutrient deprivation and ribosome dysfunction [58-61]. In the presence of such stress stimuli, p53 becomes post-translationally modified, resulting in displacement of negative regulators of p53 as highlighted below, and consequent p53 stabilization and activation. Upon activation, p53 can trigger an exquisitely complex anti-proliferative program, promoting apoptosis or cellular senescence. Thus, in cells exposed to potent stress signals, p53 drives irreversible programs of apoptosis or senescence to cull irreparably damaged or neoplastic cells. Alternatively, under conditions of low-level stress, p53 elicits protective, pro-survival responses, such as temporary cell-cycle arrest, DNA repair, and antioxidant protein production, to maintain genome integrity and viability in cells that sustain limited, reparable damage. The temporary cell-cycle arrest response is particularly well studied in the context of DNA-damage and is envisaged to allow cells an opportunity to repair DNA damage before progression through the cell cycle. This minimizes the propagation of potentially deleterious mutations – a role that led p53 to be named the "guardian of the genome." P53 drives these responses primarily by serving as a transcriptional activator of a panoply of target genes that induces programs of gene expression important for each p53 response, although transactivation-independent activities ascribed to p53 also play a role **(Figure 1.12)**.

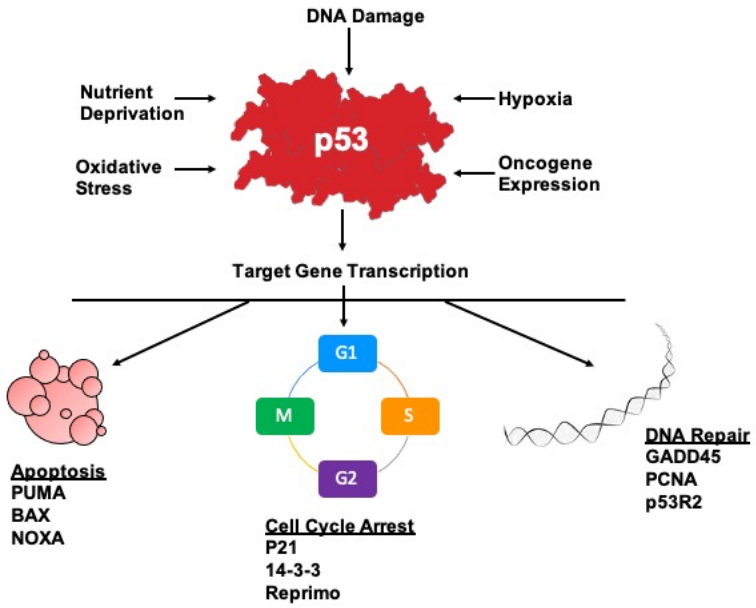


Figure 1.12. p53 Regulates Target Genes in Response to Stress Signals. P53 restricts tumor development by serving as a sensor of cellular stress responding to diverse signals, including DNA damage, hypoxia and nutrient deprivation, limiting the propagation of cells under these conditions.

The notorious p53 is a transcriptional activator

As described earlier, p53 is a cellular stress sensor, responding to such signals as DNA damage, hypoxia and hyperproliferation by inducing the cells to undergo cell cycle arrest or apoptosis. The exact cell fate specified by p53 activation is dictated by cell type, environmental milieu, and the nature of the stress. P53-triggered apoptosis involves the transcriptional induction of components of both the extrinsic and intrinsic death pathways, including *BAX*, *FAS*, *NOXA* and *PUMA*, among others, which collaboratively promote cell death [58-61]. In other cases, p53 responds to potent stress by inducing cellular senescence through transcriptional activation of target genes such as *p21*, *PAI1* and *PML.* Under conditions of lower levels of stress, when repair is possible, p53 engages a temporary program of cell-cycle arrest and DNA repair to allow cells to pause and repair any damage incurred, thereby limiting the propagation of oncogenic mutations. Notably, the tumor- suppressor function of p53 extends beyond the capacity to trigger cell-cycle arrest and apoptosis, and novel activities that impact tumor suppression are perpetually emerging, including the regulation of metabolism, autophagy, and the oxidative status of the cell. For example, the role for p53 as ‘guardian of the genome’ extends further to the maintenance of genomic stability at the chromosomal level, by limiting the accrual of

aneuploid cells. Another protective, pro- survival mechanism is the capacity of p53 to upregulate the expression of antioxidant genes, such as sestrins 1 and 2 (*SESN1* and *SESN2*, respectively), *GPX1* and *TIGAR*, which suppress the accumulation of reactive oxygen species, thereby maintaining genomic integrity [58]. Given the plethora of target genes that p53 regulates, the ability of p53 to control many cellular processes is anticipated, and many different gene expression programs presumably underscores p53's potent tumor suppressor activity. Furthermore, taken in sum, these findings attests to the centrality of p53 as a major mainstay in the body's built-in anticancer defense mechanisms.

p53 has protein domains similar to other transcriptional activators

Similar to other transcription factors, p53 has a modular protein domain structure **(Figure 1.13).** The N-terminus of p53 comprises two TADs, AD1 and AD2, which span amino acid residues 1-40 and 40-61, respectively. These domains can independently enhance transcription of p53 target genes by recruiting histone-modifying enzymes, components of the basal transcriptional machinery, and coactivator complexes, such as SAGA (SPT-ADA-GCN5 acetylase) and Mediator. C- terminal to the transactivation domains, between residues 62-95, lies the proline-rich region (PRR), which was originally proposed to participate in protein-protein interactions. The central core of p53 spanning residues 100-300 follows the PRR and comprises the DNA-binding domain that is responsible for sequence-specific DNA binding. The p53 protein binds to its response elements as a tetramer, the formation of which relies on a discrete tetramerization domain (Tet) comprising residues 325-356. Finally, p53 contains a basic, lysine-rich domain at the extreme C-terminus, between residues 363-393. This basic region binds DNA in a non- sequence-specific and undergoes extensive post-translational modifications that further modulates p53 stabilization and sequence-specific DNA binding [58-63].

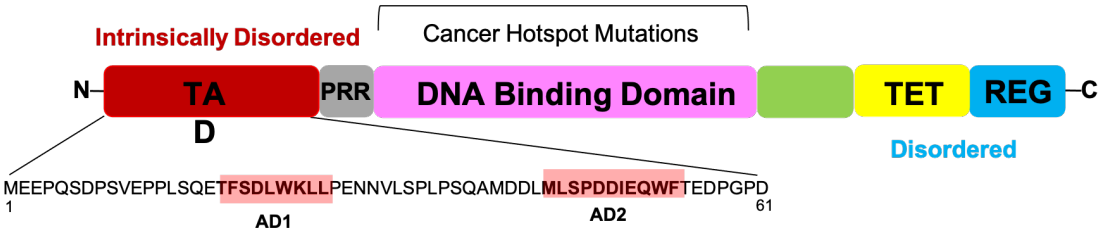


Figure 1.13. p53 Domain Architecture. Domain organization of full-length p53 consisting of an N-terminal transcriptional activation domain (TAD), followed by a proline rich region, the centrally located DNA binding domain, the tetramerization domain, and the c-terminal regulatory domain. The location of the AD1 and AD2 motifs are outlined in red on the amino acid sequence for the p53 TAD.

The DNA-binding domain

Genetic and biochemical studies alongside, structural studies have revealed molecular details of individual components of the p53, such as the DNA-binding domain and the tetramerization domain. These structures have help lay the framework for our current mechanistic understanding of p53 inactivation in cancer. Most cancer associated *p53* mutations are missense mutations in the DNA binding domain and incapacitate binding, highlighting the key importance of DNA binding for p53-mediated tumor suppression. These, tumor-derived *p53* mutations either alter residues that are essential for direct contact with p53 response elements (contact mutants) or impair proper folding of the domain (structural mutants). The six most common p53 amino acid residues altered in cancer – known as 'hotspots' – are Arg-175, Gly-245, Arg-248, Arg-249, Arg-273 and Arg-282. In addition to disrupting DNA binding, these mutations can confer gain-of-function capabilities on p53 and have been linked to increased invasiveness and metastasis of tumors. [63-65]

A multitude of structures of the central domain of p53 (p53C) have been determined revealing that p53C consists of an immunoglobulin-like β-sandwich that provides the basic scaffold for the DNA-binding surface **(Figure1.14)** [63-68]. This surface can be subdivided into two structural elements that bind to the minor groove and major groove of target DNA, respectively. The first structural element is a loop-sheet-helix motif, which docks to the DNA major groove, includes loop L1, β-strands S2 and S2′, parts of the extended β-strand S10, and the C-terminal helix. The other half of the DNA-binding surface is formed by two large loops (L2 and L3), which are stabilized by a zinc ion. The zinc ion is tetrahedrally coordinated by a histidine and three cysteine side chains (Cys-176, His-179, Cys-238, and Cys-242). Mutants that lead to the loss of binding to the Zinc

ion results in a significant decrease in thermodynamic stability, increases aggregation tendencies, and is accompanied by structural fluctuations in neighboring loops that cause loss of DNA-binding specificity. Further biochemical investigations have demonstrated that the human p53 core domain has low intrinsic thermodynamic stability with several lines of evidence suggesting that this may be the result of an adaptive evolutionary process with implications for protein turnover and binding to partner proteins. For instance, low thermodynamic stability and reduced half-lives may allow for rapid fluctuations between folded and unfolded states, which could provide an additional layer of regulation of functionally active cellular protein levels, on top of the specific degradation pathways involving ubiquitination and subsequent proteasomal degradation. Low intrinsic stability may also be directly linked with the structural plasticity required to facilitate binding to different partner proteins. Another important consideration is this low intrinsic stability of the core domain can explain the profound susceptibility of human p53 to deleterious mutations previously described and the subsequent cancer development. [63-68]

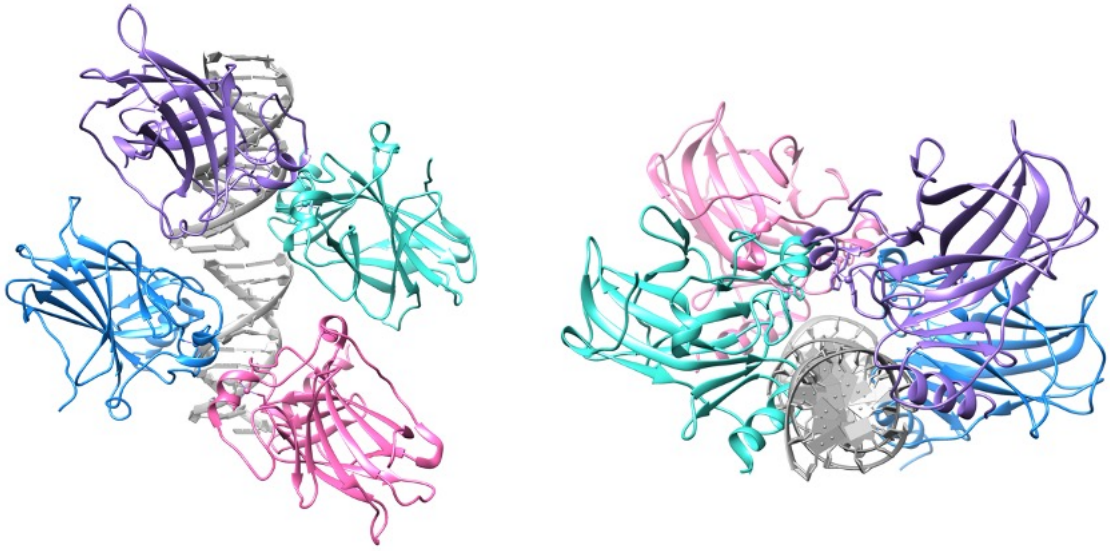


Figure 1.14. Structures of the p53 DNA Binding Domain with DNA. The crystal structure of four p53 core-domain molecules self-assemble on two B-DNA half-sites to form a tetramer.

Structural Basis of Sequence Specific DNA-Binding

The architecture of p53C tetramers bound to DNA was characterized using X-ray crystallography of p53C in complex with duplex dodecamer DNAs containing different half-site motifs. The p53 tetramer binds to sequence-specific p53 response elements, which are classically defined as two DNA half sites of RRRCWWGYYY (where R is a purine, W is adenine or thymine and Y is a pyrimidine) with a spacer of 0-13 base pairs between half sites. Two core domains bind to a half- site, forming a symmetrical dimer stabilized by base-stacking and protein-protein interactions. The symmetrical dimer forms a relatively small, self-complementary core domain interface which includes Pro-177, His-178, Arg-181, Met-243, and Gly-244. Additionally, there are key residues in the p53C-DNA interface that make direct contact with a DNA half-site (Lys-120, Ser-241, Arg-248, Arg- 273, Ala-276, Cys-277, and Arg-280) with other residues Glu-171 Arg-249 and Asp-281 forming supplementary interactions. The L3 loop (mentioned above) binds to the DNA minor groove via Arg-248, which makes either direct or water-mediated contacts with the DNA backbone. The guanidinium group of Arg-249 is essential for stabilizing the hairpin conformation of loop L3 via a network of hydrogen bonds and a salt bridge with Glu-171, thus positioning Arg-248 for DNA binding. The integral structural role of Arg-249 is emphasized by the adverse effects of the R249S cancer mutant. This mutation comprises the structural integrity of the L3 region, yielding a highly flexible structural element that favors nonnative conformations and results in abrogated DNA binding. The side chains of Arg-273 and Ser-241 makes essential contacts with the phosphate backbone at the center of the half-site. It forms a salt bridge with Asp-281, which in turn contacts DNA indirectly via a water-mediated contact. In the major groove region, Arg-280 makes invariant contacts with the conserved guanine base at position seven of the half-site, whereas contacts made by Lys-120, Ala-276, and Cys-277 from the L1 loop are sequence specific and thus modulate p53's differential target gene recognition [63, 65, 68].

Structural Insights of the Tetramerization Domain

Full-length p53 reversibly forms tetramers via a tetramerization domain in the C-terminal region of the protein (residues 325–356). The tetrameric structure of the domain, as revealed by X-ray crystallography and in solution by NMR shows the monomeric tetramerization domain consists of a short β-strand and an α-helix linked by a sharp turn

facilitated by a conserved glycine residue (Gly-334). Two monomers assemble to form a primary dimer, which is stabilized via an antiparallel intermolecular β-sheet and antiparallel helix packing with three residues (Leu-330, Ile-332, and Phe-341) forming the central hydrophobic core of this dimer. Two such dimers then associate through their helices to form a four-helix bundle tetramer. The tetramer interface is stabilized largely by hydrophobic interactions and replacement of either one of two key hydrophobic residues (Leu-344 and Leu-348) is sufficient to shift the oligomerization state toward the formation of stable dimers [63, 65, 68].

Precisely "Ordered" Transcriptional Activation through the Intrinsically "Disordered" TAD of p53

As described above, the p53 tumor suppressor is a transcriptional activator with discrete domains that participate in sequence specific DNA binding, tetramerization, and transcriptional activation. Since its original description as a tumor suppressor, many investigations have delved into the intricacies of this function seeking to understand the molecular underpinnings of p53's potent tumor suppressor activity. Although p53 was known to bind DNA through its highly conserved DNA binding domain, it was only the analysis of the composition of p53's amino terminus that revealed a significant clue into its character as a transcription factor. Notably, the amino terminus of p53 is highly acidic, a feature bearing comparison with other transcription factors. In addition, p53's amino terminus also carries a PRR, reminiscent of other activators. Definitive evidence establishing p53's role as an activator with transactivation sequences within its amino terminal residues 1-73 came from experiments showing that fusion of this region to a GAL4 DBD confers activation on a reporter carrying GAL4 binding sites. Shortly, after studies that functionally defined p53's TAD, additional work showed that p53TAD is bipartite, with two interaction motifs termed AD1 and AD2) within its N-terminus that rely on specific bulky hydrophobic residues (L22 and W23 in AD1 and W53 and F54 in AD2) for transcriptional activity. [60, 62, 63, 66, 67]

Further inquiries geared towards acquiring a deeper comprehension of the functional role of the p53 TAD employed a variety of structural biology approaches to examine both AD1 and AD2 in isolation and complexed with different interaction partners. As with other transcription activators, it was postulated that these two motifs in isolation

would be unstructured, adopting a defined structure only when in complex with a binding partner. Indeed, further experimental inquiry confirmed that the p53 TAD and its subdomains AD1 and AD2 are natively unfolded in the absence of binding partners and are easily digested by proteases, indicative of a loose structure. Circular dichroism (CD) and NMR analysis on full-length p53 as well as purified TADs in isolation (residues 1-93) revealed that p53 TADs are largely devoid of secondary and tertiary structures under. However, some NMR analyses of p53 residues 1-73 suggested that although the p53 TADs lack tertiary structure, they do contain some secondary structure. Specifically, an amphipathic α-helix is present between Thr-18 and Leu-26 and two nascent amphipathic turns, between Met-40 and Met-44 (turn 1) and Asp-48 and Trp-53 (turn 2). The significance of these structural elements is bolstered by the known importance of these residues for TAD function mentioned above. [62, 67]

Following the initial characterization of the p53 TAD in isolation, a number of additional studies examined the p53 TAD1 or TAD2 regions in complex with other proteins, the first of which was the p53 TAD1 domain complex with MDM2. MDM2 is a negative regulator of p53 that inhibits p53's transactivation potential both by binding and concealing the p53 TAD and then targeting p53 for ubiquitin-mediated proteolysis. Mdm2 directs p53 proteasomal degradation by ubiquitinating multiple lysines in the extreme carboxyl terminus of p53. [69]

The structure of the p53 TADs was initially solved by X-ray crystallography with minimal regions of both Mdm2 and p53 AD1 (amino acids 15-29) **(Figure 1.15 A).** MDM2 was found to form a deep hydrophobic cleft into which the p53 peptide inserts as an amphipathic α-helix. The primary contacts for the MDM2-p53 interaction is mediated by three highly conserved hydrophobic residues in human p53: Phe-19, Trp-23, and Leu-26. NMR analyses of a p53 peptide comprising amino acids 17-24 also highlighted the critical role of residues Phe19 and Trp23 as well as Leu22 in the p53-MDM2 interaction. [69] Another study which focuses on MDMX, a second negative regulator of p53, and the p53 AD1, reveals a similar binding pose to p53-Mdm2 interaction **(Figure 1.15 B).** [70] The crystal structure shows hydrophobic residues Phe-19, Trp-23, and Leu-26 of the p53 TAD1 forming once again the primary contact surface of the p53-binding domain of MDMX. The TAD2 of p53TAD Interacts with MDMX in a Similar Manner to MDM2. Taken

together, these studies among others show common mechanism typical of molecular recognition motif binding partner proteins often which entails disorder-to-order transitions.

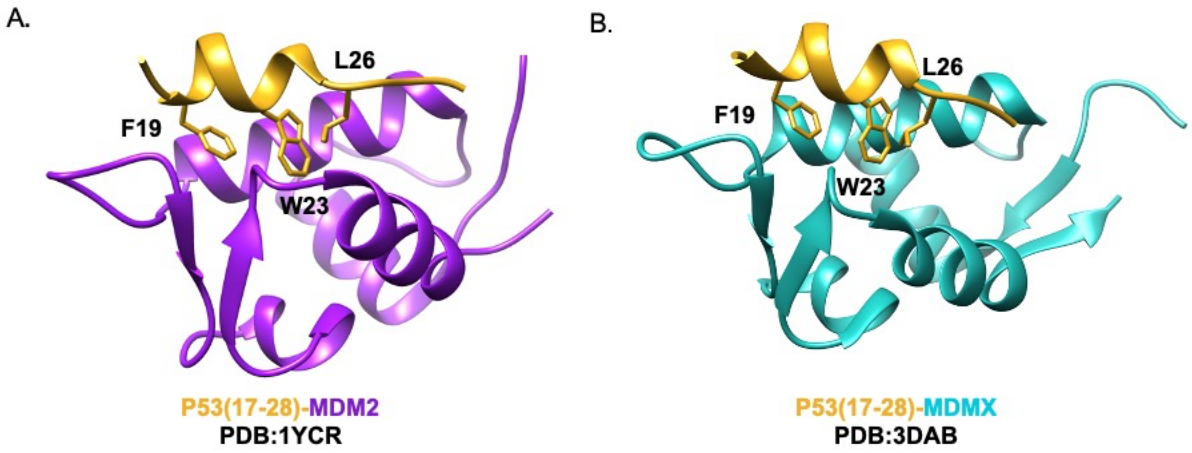


Figure 1.15. Structures of the p53 TAD with MDM2 and MDMX. The Crystal structures of the p53(17-28) N-terminal (gold) bound to **(A)** MDM2 (purple) and **(B)** MDMX (cyan) show the TAD bound as an amphipathic helix to a shallow surface cleft in each protein. The binding pose and the primary contacts (F19, W23, and L26) between the two structures remain conserved.

1.8 CBP and p53: Two Stars Center Stage in Transcriptional Regulation

As outlined above, transcriptional activators stimulate gene transcription through a series of precisely defined steps. After binding to specific DNA sites via their sequence specific DBDs, transcriptional activators must then open the adjacent chromatin and recruit the transcriptional machinery to promote RNA synthesis. Typically, through their TADs, transcriptional activators induce binding of a variety of proteins engaged in different steps of transcription, including histone modification, chromatin remodeling, and transcriptional initiation and elongation. Accordingly, the amino-terminal p53 TAD interacts with proteins involved in different steps of transcription including GTFs, RNA polymerase, and other members of the general transcriptional machinery. However, before p53 can recruit and activate these accessory proteins, p53 must first recruit coactivator proteins to DNA. CBP and p300 represent important cellular partners requisite for the activity of p53 as a tumor suppressor. In response to stress, CBP and p300 activate transcription of p53-regulated stress response genes and help stabilize p53 against ubiquitin-mediated degradation. Biological studies demonstrate that in unstressed cells, p53 is maintained at low levels through tightly coordinated interactions between

CBP or p300 and the E3 ubiquitin ligase HDM2. Following genotoxic stress, multi-site phosphorylation of the disordered p53 activation domain weakens binding to HDM2, resulting in stabilization of p53 and further enhancement of p53 binding to CBP and p300, acetylation of the disordered C-terminal regulatory region, and presumably transcriptional activation of p53 stress response genes.[66, 71] Despite the elucidation of the critical biological role of the CBP/P300 and p53 interaction the full structural basis for this interaction has yet to be elucidated. As discussed later in Chapter Two of this dissertation, this is due the complexity resulting from the larger size and inherent flexibility of CBP and p300 and p53 in addition to the intrinsic conformational heterogeneity associated with complex formation in turn precluding high resolution structure determinations in the context of full-length proteins.

1.9 Characterizing the Dynamic and Disordered: Tools at Hand for the Structural Biologist

Cryo-EM is an Important Structural Tool to Characterize Large and Dynamic Macromolecular Assemblies

Defining the molecular mechanisms underpinning biological processes necessitates the structural characterization of macromolecules and macromolecular assemblies. Traditionally, nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography are the preferred techniques for determining structures of proteins and protein complexes at atomic resolution. NMR has the advantage of being able to capture an ensemble of conformations, thus potentially allowing its users to identify both conformational changes in addition to flexible arrangements in solutions. However, the size of the sample that can be investigated by this technique is largely limited to macromolecules smaller than 50 kDa when applying conventional NMR methodologies. X-ray crystallography has also been largely successful at achieving atomic resolution both historically and presently. In this method structures are determined from the diffraction patterns of well-ordered three-dimensional crystals. However, growing well-ordered crystals in sufficient quantity and quality proves to be an immensely challenging process, especially for large and conformationally dynamic proteins and protein assemblies.

Rising in the ranks among the arsenal of high-resolution structural biology techniques is single particle electron microscopy. Unique in its ability to characterize not only large but structurally and conformationally flexible macromolecular machines, single

particle EM offers a complementary structural approach to techniques like NMR and X-ray crystallography without the need for large quantities of material or the formation of well-ordered crystals. Furthermore, recent technological breakthroughs such as direct electron detectors, and advances in image acquisition and image processing have revolutionized the capabilities of single particle EM allowing for near atomic or atomic resolution structure determination of some of the most challenging protein targets in turn shifting the focus away from the technical aspects onto the types of biological questions that can be answered. Nowhere has the promise of cryo-EM to deliver remarkable structural and mechanistic detail related to the function of large and flexible macromolecular assemblies been more apparent than in the transcription field as illustrated recent structures of Mediator and the Pre-initiation complexes. Given that single particle EM was utilized for a major portion of the studies described in this thesis a brief overview of this methodology is highlighted below.

An Overview of Electron Microscopy

Transmission electron microscopy is rapidly becoming a primary tool to answer important biological questions and allows for the study of highly dynamic proteins to understand not only the overall architecture but to derive information addressing mechanistic questions underlying macromolecular function. In particular single particle TEM allows for its end user to obtain high resolution images of protein complexes at or near atomic resolution. Single particle EM involves collecting large numbers of 2D projections of protein particles and using computational methods to sort the individual particle projections into classes based on their orientation parameters. Ultimately, the 2D particle projections can be used to generate a three-dimensional structure. There are two primary methods for single particle EM (1) negative stain electron microscopy and (2) cryo-electron microscopy, both of which rely on detecting scattered electrons after they are funneled through an intricate arrangement of electromagnetic lenses in the microscope through the biological specimen of interest. The standard workflow for cryoEM is shown below. **(Figure 1.16)**

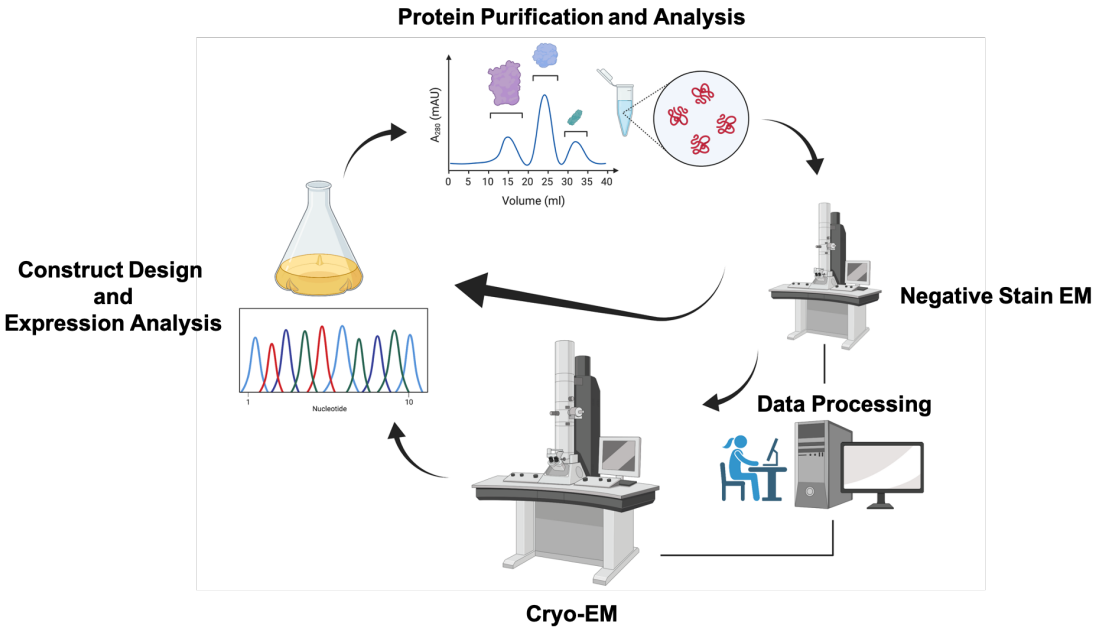


Figure 1.16. Workflow for Cryo-EM. Following purification, initial negative stain EM screening serves as a quality check regarding the suitability of a biological specimen. Heterogeneous or unstable samples as determined by negative stain EM should lead to further design of constructs, optimization of expression conditions, and purification schemes. Following initial characterization by cryo-EM additional protein chemistry can be performed to gain additional structure insights under varying biochemical states. Computational analyses is conducted throughout the EM workflow.

Negative Stain Electron Microscopy

Negative stain EM involves a protein sample being adsorbed to a continuous carbon substrate over a copper support grid. The protein specimen is then embedded and dried in a layer of heavy metal salt solution. The areas occupied by the specimen on the grid are relatively electron transparent and exclude the stain. The strong differential electron scattering between the “electron transparent” biological specimen and the electron dense surrounding negative stain results in high contrast images. These high contrast images can be used to rapidly evaluate sample characteristics such as homogeneity and oligomeric state and when coupled with 2D classification can give structural informational about the outer envelope and overall architecture of a molecule. The resolution of negative stain is limited to approximately 20 Å. Negative stain limits in resolution are due to three major factors (1) protein collapse (2) preferred orientation and (3) stain granule size. Protein collapse is a consequence of the sample preparation that induces specimen dehydration in the vacuum of the microscope. Negatively stained samples often rest on grids in preferred orientations leading to a biased view and an absence of structural information for additional orientations, hindering efforts for 3D analysis. Typically stain

granules are around 15-20 Å further limiting the resolution of this technique. Lastly, weakly binding complexes may also be disrupted and dissociate during the staining process. Notwithstanding these limitations, in favorable cases negative stain analysis can serve as an invaluable tool for obtaining low resolution models of the general architecture of difficult to study proteins. [72-74]

Cryo-EM

In single particle cryo-EM no crystals of a macromolecule are required, and only small amounts of sample are needed in contrast to X-ray crystallography. Furthermore, unlike the typical constraints imposed on conventional NMR methods, cryo-EM microscopy is well poised for the analysis of large and flexible macromolecules. In cryo-EM, small amount of aqueous sample is applied to a holy carbon EM grid which is then rapidly plunge-frozen into liquid ethane. The protein molecules frozen in the thin layer of vitreous ice are captured in multiple orientations to orient allowing for the presence of a number of different views. This makes cryo-EM the preferred single particle EM choice for high resolution structure determination because protein samples are directly imaged without stain, are protected from dehydration in the TEM vacuum, and maintain the specimen under near-native conditions, circumventing the resolution limitations imposed in negative stain EM. Probably one of the most exciting aspects of cryo-EM microscopy is that now new powerful image classification algorithms can facilitate data sorting to not only enable the determination of 3D maps with higher resolution, but also allows for the exploration of the multiple conformational and compositional states that are often present in a single sample. This simultaneous characterization of both structure and dynamics creates opportunities to illuminate the inner workings of macromolecular complexes with a level of detail we have never seen before for a richness of biological insight.

1.10 Dissertation Summary

Imaging and Imagining the Impossible

Currently there is no complete structural picture of how p300 and CBP interact with activators such as p53, although structures of individual components of CBP/p300, as well as structures of individual domains of CBP/p300 in complex with specific regions of p53, have informed model construction. By employing a multipronged structural approach including single particle electron microscopy techniques, the study outlined here will

provide snapshots of an intact activator-coactivator complex. These structural data allow for new mechanistic insights into the molecular recognition principles through which p53 recruits CBP and p300 to activate transcription, testing the hypothesis that p53 binding to CBP or p300 leads to large architectural changes.

1.11 References

1. Hahn, S., *Structure and mechanism of the RNA polymerase II transcription machinery.* Nat Struct Mol Biol, 2004. **11**(5): p. 394-403.
2. Schier, A.C. and D.J. Taatjes, *Structure and mechanism of the RNA polymerase II transcription machinery.* Genes Dev, 2020. **34**(7-8): p. 465-488.
3. Nevado, J., et al., *Transcriptional activation by artificial recruitment in mammalian cells.* Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2674-7.
4. Ptashne, M. and A. Gann, *Transcriptional activation by recruitment.* Nature, 1997. **386**(6625): p. 569-77.
5. Thomas, M.C. and C.-M. Chiang, *The General Transcription Machinery and General Cofactors.* Critical Reviews in Biochemistry and Molecular Biology, 2006. **41**(3): p. 105-178.
6. Sauer, R.T., *Scissors and helical forks.* Nature, 1990. **347**(6293): p. 514-515.
7. Steitz, T.A., *Structural studies of protein–nucleic acid interaction: the sources of sequence-specific binding.* Quarterly Reviews of Biophysics, 1990. **23**(3): p. 205-280.
8. Lee, S.H. and M. Hannink, *Molecular mechanisms that regulate transcription factor localization suggest new targets for drug development.* Adv Drug Deliv Rev, 2003. **55**(6): p. 717-31.
9. Ma, J., *Transcriptional activators and activation mechanisms.* Protein Cell, 2011. **2**(11): p. 879-88.
10. Morishita, R., *Transcription factor as molecular targets: is transcription factor decoy a novel drug?* Curr Drug Targets, 2003. **4**(8): p. 2 p before 599.
11. Liu, J., et al., *Intrinsic Disorder in Transcription Factors†.* Biochemistry, 2006. **45**(22): p. 6873-6888.
12. Papavassiliou, K.A. and A.G. Papavassiliou, *Transcription Factor Drug Targets.* J Cell Biochem, 2016. **117**(12): p. 2693-2696.
13. Staby, L., et al., *Eukaryotic transcription factors: paradigms of protein intrinsic disorder.* Biochemical Journal, 2017. **474**(15): p. 2509-2532.
14. Dunker, A.K. and V.N. Uversky, *Drugs for 'protein clouds': targeting intrinsically disordered transcription factors.* Curr Opin Pharmacol, 2010. **10**(6): p. 782-8.
15. Dyson, H.J. and P.E. Wright, *Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300.* J Biol Chem, 2016. **291**(13): p. 6714-22.
16. Wang, F., C.B. Marshall, and M. Ikura, *Transcriptional/epigenetic regulator CBP/p300 in tumorigenesis: structural and functional versatility in target recognition.* Cell Mol Life Sci, 2013. **70**(21): p. 3989-4008.
17. Jeronimo, C. and F. Robert, *The Mediator Complex: At the Nexus of RNA Polymerase II Transcription.* Trends Cell Biol, 2017. **27**(10): p. 765-783.
18. Robinson, P.J., et al., *Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex.* Cell, 2016. **166**(6): p. 1411-1422 e16.
19. Poss, Z.C., C.C. Ebmeier, and D.J. Taatjes, *The Mediator complex and transcription regulation.* Crit Rev Biochem Mol Biol, 2013. **48**(6): p. 575-608.
20. Soutourina, J., *Transcription regulation by the Mediator complex.* Nat Rev Mol Cell Biol, 2018. **19**(4): p. 262-274.
21. Thakur, J.K., A. Yadav, and G. Yadav, *Molecular recognition by the KIX domain and its role in gene regulation.* Nucleic Acids Res, 2014. **42**(4): p. 2112-25.

22. Lill, N.L., et al., *Binding and modulation of p53 by p300/CBP coactivators.* Nature, 1997. **387**(6635): p. 823-827.
23. Yu, Q., et al., *Global Conformational Selection and Local Induced Fit for the Recognition between Intrinsic Disordered p53 and CBP.* PLoS ONE, 2013. **8**(3): p. e59627.
24. Arai, M., J.C. Ferreon, and P.E. Wright, *Quantitative Analysis of Multisite Protein–Ligand Interactions by NMR: Binding of Intrinsically Disordered p53 Transactivation Subdomains with the TAZ2 Domain of CBP.* Journal of the American Chemical Society, 2012. **134**(8): p. 3792-3803.
25. Nyqvist, I., E. Andersson, and J. Dogan, *Role of Conformational Entropy in Molecular Recognition by TAZ1 of CBP.* The Journal of Physical Chemistry B, 2019. **123**(13): p. 2882-2888.
26. Law, S.M., et al., *Prepaying the entropic cost for allosteric regulation in KIX.* Proceedings of the National Academy of Sciences, 2014. **111**(33): p. 12067-12072.
27. Attar, N. and S.K. Kurdistani, *Exploitation of EP300 and CREBBP Lysine Acetyltransferases by Cancer.* Cold Spring Harb Perspect Med, 2017. **7**(3).
28. Darnell, J.E., *Transcription factors as targets for cancer therapy.* Nature Reviews Cancer, 2002. **2**(10): p. 740-749.
29. Højfeldt, J.W., A.R. Van Dyke, and A.K. Mapp, *Transforming ligands into transcriptional regulators: building blocks for bifunctional molecules.* Chemical Society Reviews, 2011. **40**(8): p. 4286.
30. Kuttler, F. and S. Mai, *c-Myc, Genomic Instability and Disease*, in *Genome and Disease*. 2006, KARGER. p. 171-190.
31. Nesbit, C.E., J.M. Tersak, and E.V. Prochownik, *MYC oncogenes and human neoplastic disease.* Oncogene, 1999. **18**(19): p. 3004-3016.
32. Allen, B.L. and D.J. Taatjes, *The Mediator complex: a central integrator of transcription.* Nature Reviews Molecular Cell Biology, 2015. **16**(3): p. 155-166.
33. Napoli, C., et al., *Unraveling framework of the ancestral Mediator complex in human diseases.* Biochimie, 2012. **94**(3): p. 579-587.
34. Dancy, B.M. and P.A. Cole, *Protein Lysine Acetylation by p300/CBP.* Chemical Reviews, 2015. **115**(6): p. 2419-2452.
35. Park, S., et al., *Role of the CBP catalytic core in intramolecular SUMOylation and control of histone H3 acetylation.* Proceedings of the National Academy of Sciences, 2017. **114**(27): p. E5335-E5342.
36. Bekesi, A., et al., *Challenges in the Structural-Functional Characterization of Multidomain, Partially Disordered Proteins CBP and p300: Preparing Native Proteins and Developing Nanobody Tools.* Methods Enzymol, 2018. **611**: p. 607-675.
37. Huang, Y., et al., *Deciphering the promiscuous interactions between intrinsically disordered transactivation domains and the KIX domain.* Proteins: Structure, Function, and Bioinformatics, 2017. **85**(11): p. 2088-2095.
38. Odoux, A., et al., *Experimental and molecular dynamics studies showed that CBP KIX mutation affects the stability of CBP:c-Myb complex.* Computational Biology and Chemistry, 2016. **62**: p. 47-59.
39. Kasper, L.H., et al., *Genetic Interaction between Mutations in c-Myb and the KIX Domains of CBP and p300 Affects Multiple Blood Cell Lineages and Influences Both Gene Activation and Repression.* PLoS ONE, 2013. **8**(12): p. e82684.

40. Lee, C.W., et al., *Mapping the interactions of the p53 transactivation domain with the KIX domain of CBP.* Biochemistry, 2009. **48**(10): p. 2115-24.
41. Gao, M., et al., *Intrinsically Disordered Transactivation Domains Bind to TAZ1 Domain of CBP via Diverse Mechanisms.* Biophysical Journal, 2019. **117**(7): p. 1301-1310.
42. Berlow, R.B., et al., *Role of Backbone Dynamics in Modulating the Interactions of Disordered Ligands with the TAZ1 Domain of the CREB-Binding Protein.* Biochemistry, 2019. **58**(10): p. 1354-1362.
43. Lindström, I., E. Andersson, and J. Dogan, *The transition state structure for binding between TAZ1 of CBP and the disordered Hif-1α CAD.* Scientific Reports, 2018. **8**(1).
44. Berlow, R.B., H.J. Dyson, and P.E. Wright, *Hypersensitive termination of the hypoxic response by a disordered protein switch.* Nature, 2017. **543**(7645): p. 447-451.
45. Miller Jenkins, L.M., et al., *Characterization of the p300 Taz2–p53 TAD2 Complex and Comparison with the p300 Taz2–p53 TAD1 Complex.* Biochemistry, 2015. **54**(11): p. 2001-2010.
46. De Guzman, R.N., et al., *Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP.* J Mol Biol, 2000. **303**(2): p. 243-53.
47. Feng, H., et al., *Structural Basis for p300 Taz2-p53 TAD1 Binding and Modulation by Phosphorylation.* Structure, 2009. **17**(2): p. 202-210.
48. Lochhead, M.R., et al., *Structural insights into TAZ2 domain–mediated CBP/p300 recruitment by transactivation domain 1 of the lymphopoietic transcription factor E2A.* Journal of Biological Chemistry, 2020. **295**(13): p. 4303-4315.
49. Kjaergaard, M., K. Teilum, and F.M. Poulsen, *Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP.* Proceedings of the National Academy of Sciences, 2010. **107**(28): p. 12535-12540.
50. Kjaergaard, M., et al., *A Folded Excited State of Ligand-Free Nuclear Coactivator Binding Domain (NCBD) Underlies Plasticity in Ligand Recognition.* Biochemistry, 2013. **52**(10): p. 1686-1693.
51. Haberz, P., et al., *Mapping the interactions of adenoviral E1A proteins with the p160 nuclear receptor coactivator binding domain of CBP.* Protein Science, 2016. **25**(12): p. 2256-2267.
52. Henry, R.A., et al., *Changing the selectivity of p300 by acetyl-CoA modulation of histone acetylation.* ACS Chem Biol, 2015. **10**(1): p. 146-56.
53. Thompson, P.R., et al., *Regulation of the p300 HAT domain via a novel activation loop.* Nat Struct Mol Biol, 2004. **11**(4): p. 308-15.
54. Bose, D.A., et al., *RNA Binding to CBP Stimulates Histone Acetylation and Transcription.* Cell, 2017. **168**(1-2): p. 135-149.e22.
55. Delvecchio, M., et al., *Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation.* Nat Struct Mol Biol, 2013. **20**(9): p. 1040-6.
56. Ramsay, R.G., A.L. Barton, and T.J. Gonda, *Targeting c-Myb expression in human disease.* Expert Opinion on Therapeutic Targets, 2003. **7**(2): p. 235-248.
57. Zor, T., et al., *Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of c-Myb.* Journal of Molecular Biology, 2004. **337**(3): p. 521-534.
58. Mills, K.D., *Tumor suppression: Putting p53 in context.* Cell Cycle, 2013. **12**(22): p. 3461-3462.

59. Kastenhuber, E.R. and S.W. Lowe, *Putting p53 in Context.* Cell, 2017. **170**(6): p. 1062-1078.
60. Brady, C.A. and L.D. Attardi, *p53 at a glance.* Journal of Cell Science, 2010. **123**(15): p. 2527-2532.
61. Mello, S.S. and L.D. Attardi, *Deciphering p53 signaling in tumor suppression.* Current Opinion in Cell Biology, 2018. **51**: p. 65-72.
62. Raj, N. and L.D. Attardi, *The Transactivation Domains of the p53 Protein.* Cold Spring Harbor Perspectives in Medicine, 2017. **7**(1): p. a026047.
63. Joerger, A.C. and A.R. Fersht, *Structural biology of the tumor suppressor p53.* Annu Rev Biochem, 2008. **77**: p. 557-82.
64. Bista, M., S.M. Freund, and A.R. Fersht, *Domain-domain interactions in full-length p53 and a specific DNA complex probed by methyl NMR spectroscopy.* Proceedings of the National Academy of Sciences, 2012. **109**(39): p. 15752-15756.
65. Weinberg, R.L., D.B. Veprintsev, and A.R. Fersht, *Cooperative Binding of Tetrameric p53 to DNA.* Journal of Molecular Biology, 2004. **341**(5): p. 1145-1159.
66. Joerger, A.C. and A.R. Fersht, *The tumor suppressor p53: from structures to drug discovery.* Cold Spring Harb Perspect Biol, 2010. **2**(6): p. a000919.
67. Wells, M., et al., *Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain.* Proceedings of the National Academy of Sciences, 2008. **105**(15): p. 5762-5767.
68. Veprintsev, D.B., et al., *Core domain interactions in full-length p53 in solution.* Proceedings of the National Academy of Sciences, 2006. **103**(7): p. 2115-2119.
69. Kussie, P.H., et al., *Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain.* Science, 1996. **274**(5289): p. 948-953.
70. Popowicz, G., A. Czarna, and T. Holak, *Structure of the human Mdmx protein bound to the p53 tumor suppressor transactivation domain.* Cell Cycle, 2008. **7**(15): p. 2441-2443.
71. Ferreon, J.C., et al., *Cooperative regulation of p53 by modulation of ternary complex formation with CBP/p300 and HDM2.* Proc Natl Acad Sci U S A, 2009. **106**(16): p. 6591-6.
72. Skiniotis, G. and D.R. Southworth, *Single-particle cryo-electron microscopy of macromolecular complexes.* Microscopy (Oxf), 2016. **65**(1): p. 9-22.
73. Ohi, M., et al., *Negative staining and image classification — powerful tools in modern electron microscopy.* Biological Procedures Online, 2004. **6**(1): p. 23-34.
74. Peisley, A. and G. Skiniotis, *2D Projection Analysis of GPCR Complexes by Negative Stain Electron Microscopy*, in *Methods in Molecular Biology*. 2015, Springer New York. p. 29

CHAPTER TWO

Three-Fifths is not a Whole: Biochemical and Biophysical Characterization of an Intact Transcriptional Activator Coactivator Complex”

Abstract

Challenge Accepted: “Scientific Fearlessness” Embraced by the Black and Brilliant

Structural disorder is pervasive in proteins that reside at the center of transcription networks in eukaryotes as is the case of p53 and its regulatory binding partners CBP and p300. p53 and the closely related coactivators p300 and CBP are all large multidomain proteins bearing considerable regions of disorder. Disorder can provide many advantages to the function and regulation of these proteins including conformational plasticity, promiscuity, and regulation via a diverse ornamentation of post-translational modifications (PTMs). However, the structural and functional analysis of large multidomain proteins containing regions of disorder like p53 and CBP/p300 present unique challenges to the structural biologist. Part of the challenge is being able to express, isolate, and purify a protein or complex, that is not only large, but is temperamental, aggregation prone, and compromised by a vulnerability to proteolytic cleavage. Thus, there is a dire need to develop better platforms to express and purify proteins of this nature.

In this chapter I describe the effective expression of full-length CBP, p300, and p53 in a eukaryotic host, followed by their successful isolation and purification to near homogeneity. I also report the first *in vitro* biochemical reconstitution of the functionally important complexes formed between the full-length master coactivators CBP and p300 and the full length tetrameric p53. Importantly, I show initial negative stain EM studies that not only reveal the overall shape of these transcriptional activator complexes but

demonstrate the conformationally dynamic nature of this system. While the resolution of the initial negative stain analyses is low for the CBP•$(p53)_4$ complexes, the studies herein support the feasibility of more detailed structural analyses utilizing cryo-EM and demonstrate the promise for structural and dynamic studies of IDPs.

2.1 Introduction

Full Measures-A Journey to Wholeness by Sifting Through the Broken Pieces

Multiple lines of evidence suggest that upon DNA damage CBP and p300 are recruited by p53 to modify chromatin and aid in transcriptional activation of p53 target genes. Key support for the critical role of CBP and p300 in the regulation of p53‘s transcriptional activation came from a number of studies including one from the laboratory of Kathleen Kelly in which a gel super shift assay was used to place the p300-p53 complex on DNA at p53-dependent promoters. In the same study, Kelly and colleagues were able to demonstrate in luciferase assays that when MEF cells lacking endogenous p53 were transfected with several p300 constructs alone the addition of these proteins had no significant effect on the levels of activation in the absence of wildtype p53. Unsurprisingly, they showed that transfection of p53 produced a marked stimulation but significant stimulation of the reporter above the levels observed with p53 alone was specifically induced by co-expression with full-length p300 (1). Following these reports, further insights revealing additional details of the CBP/p300-p53 interaction in p53 dependent transcription came from binding assays highlighting that the KIX, TAZ1, TAZ2, and NCBD domains of CBP and p300 all bind the p53 tetramer albeit with varying affinities (16). Lastly, detailed structure function analysis have been performed for these more “ordered” domains of CBP and p300 and the complexes they form with the isolated p53 TAD. These studies reveal not only the structural versatility in the binding poise as each complex exhibits unique tertiary conformations but also provides important insight into the molecular recognition principles that guide interactions between p53 and CBP and p300. Collectively, these findings have led to a model where each of the four N-terminal transcriptional activation domains of an active p53 tetramer bind to the four separate activator binding domains of a single CBP or p300 molecule. In turn, this stabilizes the p53 tetramer against proteasomal degradation that would otherwise occur and enhances p53-mediated transcription **(Figure 2.1).**

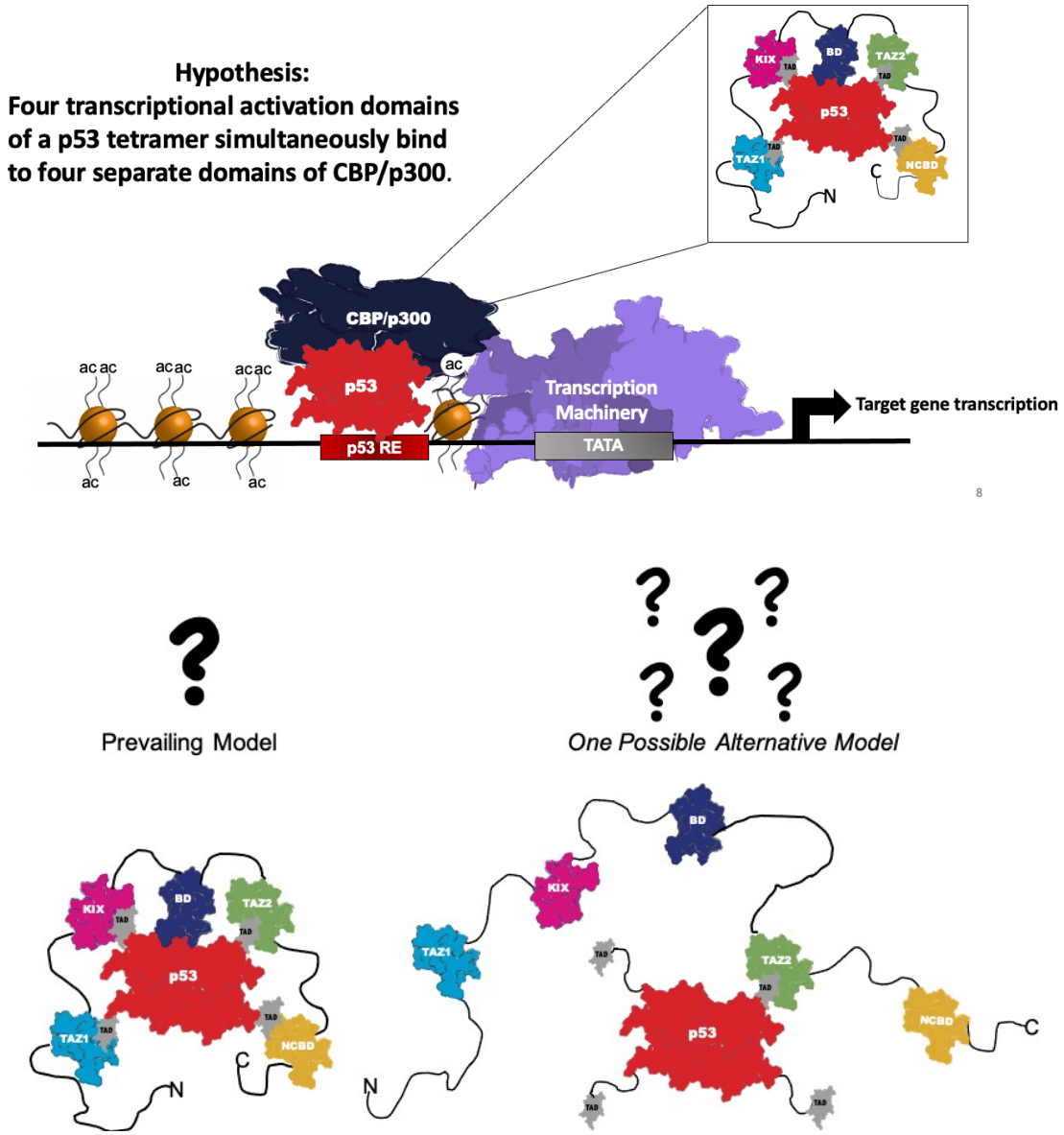


Figure 2.1. Prevailing and Alternative Models for the p53-CBP /p300 Interaction. In the prevailing model for the CBP/p300-p53 interaction four transcriptional activation domains of tetrameric p53 interact with the binding sites of CBP/p300. The bromodomain of CBP/p300 may further stabilize this complex by interacting with acetylated lysine 382 of p53. For simplicity, the HAT, PHD, and ZZ domains of CBP/p300 are not shown. This model however does not take into account the other possible stoichiometric arrangements that could occur during higher-order complex assembly.

While the aforementioned investigations have advanced our knowledge on this system and lend support to the current model nonetheless one cannot overlook that large modular proteins do not function as isolated domains and that all the structures described with “minimal” TADs of p53 and an isolated ABD of CBP or p300 do not capture the complete picture. Specifically, these structural escapades fail to capture the synergy

between the ordered and disordered elements that presumably mediate proper protein function within CBP, p53, p300, and within the complexes they form. Therefore, many lingering questions remain about the mechanism of these intriguingly complex systems that can only be answered through the detailed structural characterizations of intact proteins. To date, however, very few studies exists on full length transcriptional activators and the complexes they form with full-length coactivators. Typically, these kinds of proteins are quite large, consisting of mixtures of independently folded domains separated by intrinsically disordered regions that are frequently the sites of a number of PTMS. Given these properties, structural studies have lagged behind on activator-coactivator complexes in part because the ability to purify the biomolecules in an intact stable manner represents a major hurdle.

This is nowhere more apparent than in the case of the functionally critical complex formed between CBP or p300 and the transcriptional activator p53. As outlined in chapter one of this dissertation, CBP and p300 are large proteins that function as transcriptional coactivators largely through use of their multiple dynamic ABD domains that are connected by regions that are predicted to bear significant structural disorder. Similar to CBP and p300, p53 exerts its function through an intricate interplay of utilizing both independently folded and intrinsically disordered functional domains. Thus, the importance of improved expression and purification strategies in addition to new biophysical methodologies for their analyses cannot be overstated for the mechanistic and structural interrogation of transcriptional systems like the one described above. In this chapter, I described the development and optimization of reliable methods for expressing and purifying CBP, p300, and p53 suitable for biochemical and structural studies. Furthermore, I show that both full-length CBP and p300 are capable of binding full-length p53 in vitro. Significantly, I provide initial characterization of the CBP, p300, and their bound complexes with p53 by negative stain EM complexes demonstrating the feasibility of the study of these macromolecular complexes for more in-depth structural work.

2.2 Results and Discussion

"Expressive" and Expressed: Expression and Purification of the Large and Intrinsically Disordered CBP, p300, and p53

The production of sufficient quantities of homologous protein is the starting point for any successful structural biology investigation. In order to obtain protein of suitable quality for structural characterization, initial efforts of my work were focused on the optimization of the expression of the full-length CBP, p300, and p53 in a suitable expression host. Although *Escherichia coli* is the most commonly employed host to produce recombinant proteins for structural studies because of its versatility and costs its use an expression system had significant drawbacks in our case. In particular, bacteria are unable to provide post-translational modifications and folding aids such as chaperones required for the generation of fully functional eukaryotic proteins. In contrast to *E. coli*, insect cells have the machinery required for proper folding, post-translational modification, and authentic processing capabilities requisite for overexpressing functional proteins. Moreover, high-level production of recombinant proteins in baculovirus-infected insect cells is rarely associated with inclusion body formation, which is commonly observed in bacterial systems. In light of these considerations, we chose to commence expression analyses in insect cells. To this end, the full-length CBP and p300 genes were cloned into baculovirus transfer vectors for use in the baculovirus expression system (Performed by Dr. Jean Lodge). Each construct was cloned in frame with various fusion tags to aid in the subsequent protein purification steps (Performed by Dr. Jean Lodge). Insect cell expression constructs for full-length CBP and p300 were also obtained from the laboratories of Dr. Shelley Berger (University of Pennsylvania) and Dr. Andy Andrews (Fox Chase Cancer Center) respectively. Additionally, a suitable expression vector for full-length MBP-tagged p53 was obtained from Dr. Daniel Southworth (University of Michigan-Formerly, University of California San Francisco -Currently).

For all constructs baculoviral stocks for protein expression were generated. Optimal expression conditions were evaluated by performing small-scale expression trials in baculovirus infected cells encompassing time course experiments where the following parameters were varied: media additives (including Fetal Bovine Serum (FBS) to protect the recombinant protein from proteolysis), growth temperature, and multiplicity of infection

(MOI). Cells were usually analyzed 24, 48, and 72 hours post infection as some stable proteins might accumulate to high levels 72 post infection while others, sensitive to degradation, will need to be collected after 24 hours or most commonly 48 hours. Given that protein expression may also depend on cell line, expression trials were conducted in two insect cell lines: Sf9 and High Five. To determine the relative levels of protein expression for each series on experiments under varied parameters, pellets from small scale cell growths were collected at their respective time points, lysed, and then assayed for protein detection. Recombinant protein was isolated by performing small batch immobilized metal affinity chromatography on the lysates and subsequent SDS-PAGE analysis.

After an extensive series of expression trials, suitable conditions were identified for the expression of one full-length CBP construct and one full-length p300 construct. Suitable expression conditions had previously been identified for p53 expression in sufficient quantities in SF9 cells (Ulla Lilienthal). As shown, in (**Figure 2.2)**, panel 2 when compared to the same infection parameters in the SF9 cell line, the highest detectable level of protein observed for full-length CBP is at the 24-hour time period post infection with the lowest screened infection volume of CBP baculovirus (30 μL per 5 ml of cell culture) in the High Five cell line. This is observed in comparison with the 48-hour time point. At the 72-hour time point, CBP expression could not be detected in either cell lines, perhaps indicative of cell lysis from viral infection or ongoing proteolysis (data not shown). Furthermore, the addition of FBS did not seem to improve protein quantities or stability in our studies. Analogous studies were carried out for the full-length p300 construct. In contrast to the CBP construct, no detectable expression was observed 24 hours post infection. Protein expression was optimal for full-length p300 in High Five cells and SF9 cells at the 48- and 72-hour time points, respectively **(Figure 2.3)**. Of note, there is considerable background for both expression of CBP and p300, potentially indicative of the presence of co-purifying proteins or degradation of the full-length constructs. Given that this background does not persist in the time course experiments with p300, this strongly suggested that the full-length p300 construct was perhaps less stable than our CBP construct and more prone to degradation. Therefore, we pursued large scale

expression for the CBP construct under the favorable expression conditions identified for our initial studies.

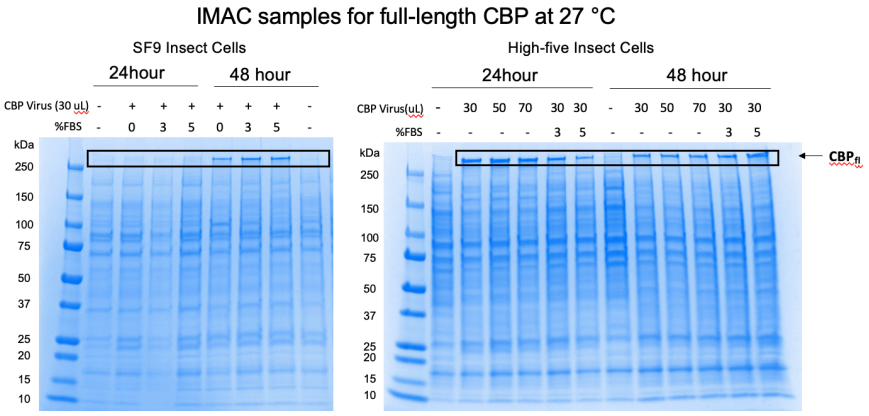


Figure 2.2 Expression Trial Analysis in Insect Cells for full-length CBP. SDS-PAGE analysis evaluating protein expression for $CBP_{fl\ in}$ SF9 insect cells (left gel) and high five cells (right gel) at 24 hours and 48 hours post infection under various conditions. The molecular weight of CBP_{fl} is ~ 265 kDa. The expression of CBP is indicated (outlined in black) in the gel at its corresponding molecular weight.

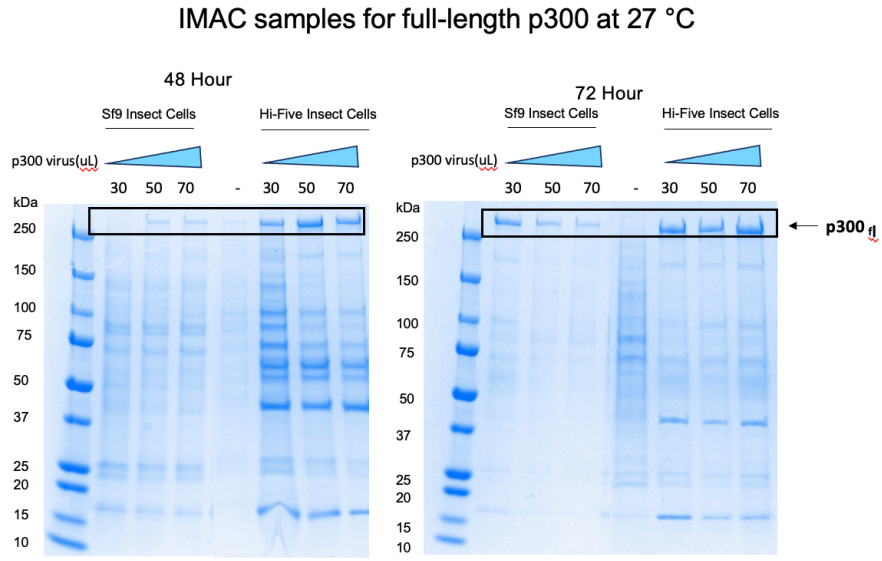


Figure 2.3 Expression Trial Analysis in Insect Cells for full-length p300. SDS-PAGE analysis evaluating protein expression for $p300_{fl}$ in SF9 insect and high five cells high at 48-hour (left gel) and 72 (right gel) hours post infection under various conditions. The molecular weight of $p300_{l}$ is ~ 265 kDa. The expression of p300 is indicated in the gel (outlined in black.) at its corresponding molecular weight.

Focused on the Outcome not the Obstacles: The Saga Continues, Interactive Rounds of Biochemistry for the Perfect Protein Specimen

Given that the success of a structural biology endeavor goes hand in hand with the ability to purify biochemical specimens to near homogeneity, the next major objective of my work was to obtain pure recombinant full-length CBP for structural characterization. Based on previous reports, a purification scheme relying on three purification steps including a metal affinity chromatography step, an ion exchange chromatography step, and a final gel filtration chromatography step was applied. For full-length CBP the size exclusion chromatography profile indicates a narrow peak distribution although analysis of the corresponding SDS-PAGE reveals some heterogeneity in the preparation likely due to contamination of transcriptional activators. (**Figure 2.4**). As stated earlier for full-length p53, a pFastBac construct for expressing MBP-tagged full-length p53 in SF9 insect cells was acquired from the laboratory of Dr. Daniel Southworth (University of Michigan, University of California San Francisco -Currently). Initial experiments show that this protein expresses well and the tetrameric form of p53 is capable of being purified to near homogeneity following a similar purification scheme as CBP **(Figure 2.5)**.

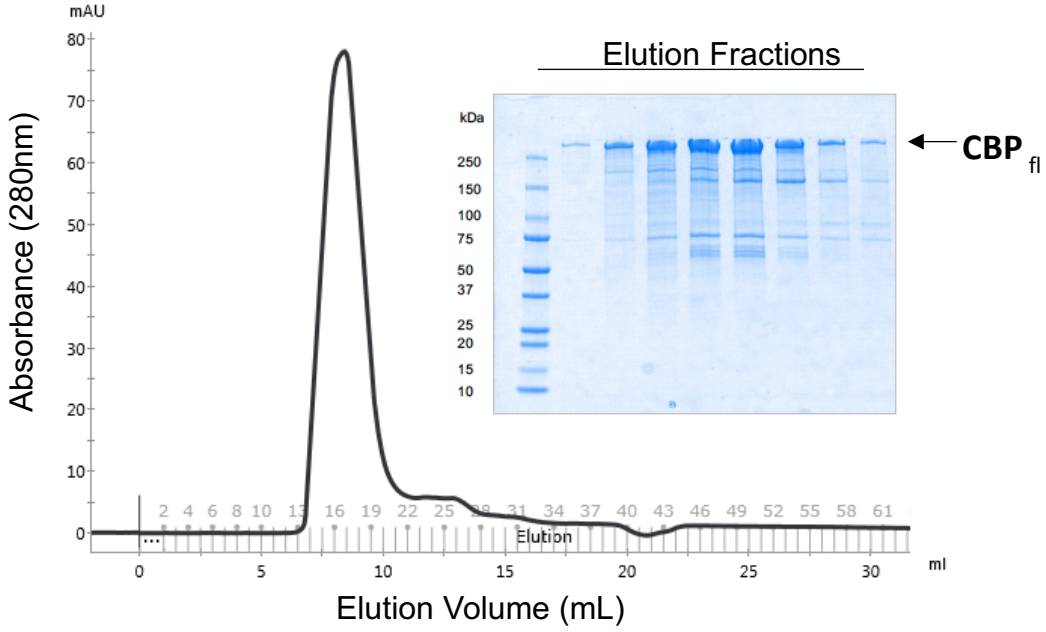


Figure 2.4 Purification of CBP$_{fl}$. Size exclusion chromatography elution profile of full-length CBP run on a Superdex 200 Increase 10/300 column and corresponding SDS-PAGE of peak fractions.

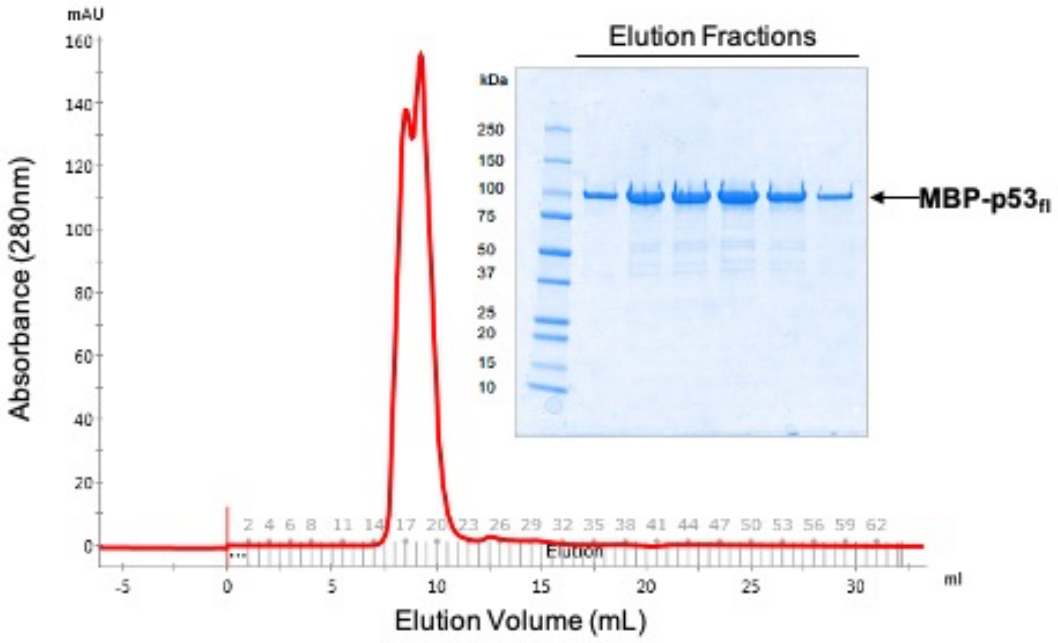


Figure 2.5 Purification of p53$_{fl}$. Size exclusion chromatography elution profile of the MBP-tagged p53 run on a Superdex 200 Increase 10/300 column and corresponding SDS-PAGE of peak fractions.

Following the successful expression and purification of CBP and p53, preparation of the binary complex between the two proteins was pursued through in vitro reconstitution. Complex formation was analyzed by size exclusion chromatography and eluted fractions were subjected to SDS-PAGE analysis to confirm the presence of both full-length CBP and full-length p53 in each fraction (**Figure 2.6**). The elution position of the complex is consistent with a molecular weight species that corresponds the formation of a binary complex formed between full-length CBP and tetrameric p53. The presence of a degree of asymmetry in the gel filtration peak may be a result of CBP co-purifying proteins from the preparation as well as the presence of p53 in differing oligomeric states when unbound to CBP. Following purification, p53, CBP, and the complex were subjected to initial negative stain electron microscopy to rapidly assess the samples quality and characteristics such as homogeneity and the oligomeric state under varying biochemical conditions. Unfortunately for all samples screened by negative stain EM significant heterogeneity was evident likely due to the presence of not only contaminants, but significant degradation of the proteins (**Figure 2.7**).

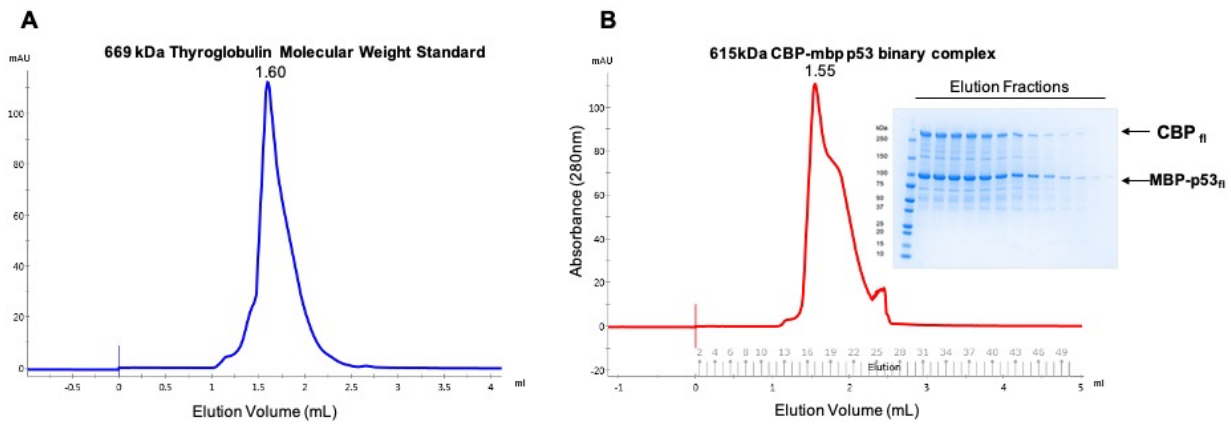


Figure 2.6. **Analytical Characterization of the CBP•$(p53)_4$ binary complex. (A)** Size exclusion chromatography elution profile of the 669kDa thyroglobulin molecular weight standard run on a Superose 6 Increase 3.2/300 analytical column. (**B.)** Size exclusion chromatography profile of the CBP-p53 binary complex run on a Superose 6 Increase 3.2/300 analytical column and corresponding SDS-PAGE of peak fractions.

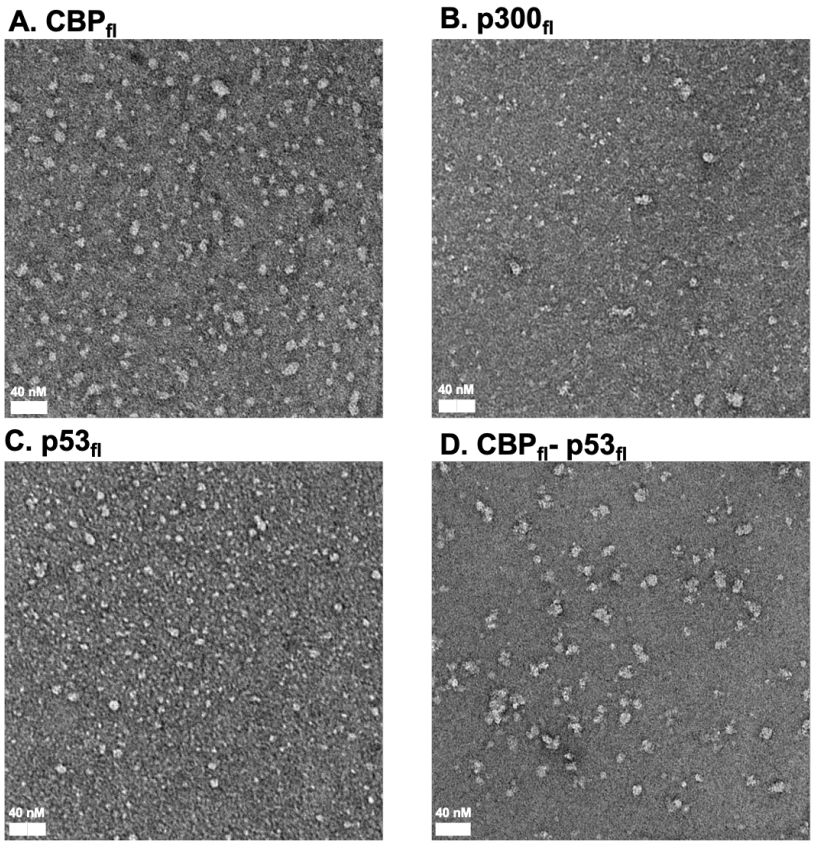


Figure 2.7. Negative Stain EM of Analysis of Purified CBP_{fl}, $p300_{fl}$, $p53_{fl}$, and CBP-$p53_{fl}$. Representative Negative Stain EM Micrographs of **(A)** CBP_{fl} (top left) **(B)**$p300_{fl}$ (top right) **(C)**$p53_{fl}$ (bottom right) and (D) the CBP $_{fl}$ -p53 $_{fl}$ binary complex (bottom left). 40nm scale bar shown on bottom right of each micrograph.

To address this challenge, further expression trials were carried out to co-express the two proteins in either SF9 or hi-five insect cells. We hypothesized that co-expression of CBP and p53 would lead to limited degradation, as well attenuate the affinity of contaminating transcription factors to bind the proteins resulting in a more homogenous biological specimen. Furthermore, it is well documented that in the absence of their interacting partners, proteins are often insoluble, improperly folded, or non-functional. Pursuant to this, additional expression trial tests were conducted in attempt to co-express CBP and p53 by co-infecting both SF9 cells and hi-five cell with virus from CBP and p53 viral stocks. Promising expression trial conditions were identified by co-infecting equal volumes of CBP and p53 (40 uL total) in the SF9 and high five cell lines after harvesting the cell pellets 48 hours post infection **(Figure 2.8)** Preparations for co-expressed proteins were scaled up and further optimization of the biochemical preparations of full-length CBP alone allowed us to isolate biological specimens suitable for EM analyses.

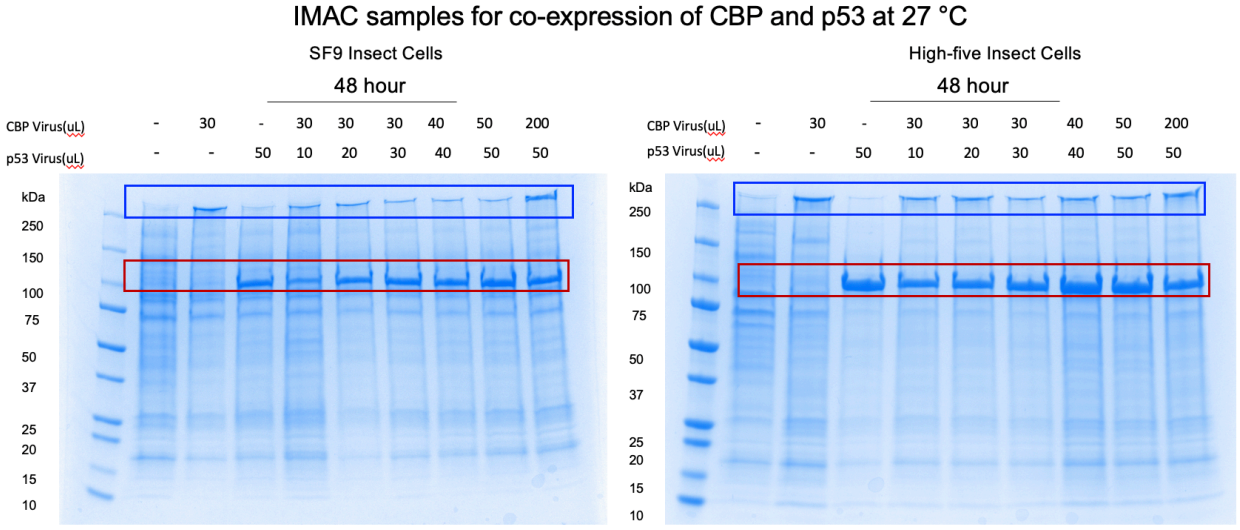


Figure 2.8 Co-expression Trial Analysis in Insect Cells for full-length CBP (CBP_{fl}) and $p53_{fl}$. SDS-PAGE analysis evaluating protein expression for CBP_{fl} and $p53_{fl}$ in SF9 insect (left gel) and high five cells (right gel) 48 hours post infection under various conditions. The molecular weight of CBP_{fl} is ~ 265 kDa and mbp-tagged monomer p53 ~87.5 kDa. The expression of CBP_{fl} (outlined in blue) and $p53_{fl}$ (outlined in red) are indicated in the gel at their corresponding molecular weights.

My Authentic Self "Fuzzy and Perfectly Functional"-Overall Architecture of CBP and its complex with p53 as revealed by Negative Stain EM

After developing a reproducible purification method to produce a sample suitable for single particle EM analysis, I next examined the sample of CBP and the CBP•$(p53)_4$ binary complex by negative stain EM. Negative stain EM was chosen because of its ability to provide information about the morphology, organization, and heterogeneity of sample in a relatively short period of time. Images of the negative stained specimen revealed a relatively homogenous, roughly spherical sample for full-length CBP with fairly uniform sizes. (**Figure 2.9**) A representative negative stain image of the CBP•$(p53)_4$ binary complex **(Figure 2.10)** also reveals a homogenous sample that shows triangular shaped particles that are suitable for initial image classification by negative stain EM and potential further examination cryo-EM. To gain further insight into the architecture of CBP and how this architecture is modulated by specific binding events we coupled our negative stain EM with single particle classification and averaging to visualize preparations consisting of the purified CBP•$(p53)_4$ and compare it with CBP alone.

We hypothesized that binding of p53 at one or more ABDs in CBP would lead to large scale global structural reorganizations in coactivator. To test this hypothesis 3,156 particles from the CBP•$(p53)_4$ images were selected and grouped into 20 classes using CryoSparc. The same was done for 1,835 particles projections of CBP. Two-dimensional (2D) classification reveals an asymmetric structural arrangement of the CBP•$(p53)_4$ binary complex with most classes from this analysis exhibiting a triangular base with small oval lobes protruding near the sides of the complex **(Figure 2.11B).** Interestingly however, a few classes are ellipsoidal and extended in appearance. The averaged particles all appear to differ slightly in their shape and overall orientations which could represent different views of the aligned particle projections or could be indicative of the presence of flexible conformers present in distinct conformations representing highly flexible particle ranging from elongated to more compact conformations. Alternatively, the differences in architecture could also be attributed to compositional and chemical heterogeneity. In comparison to the CBP•$(p53)_4$ binary complex, 2D classes of CBP show a round circular

structure, supporting the possibility of global architecture arrangements within the CBP upon activator binding **(Figure 2.11A)**.

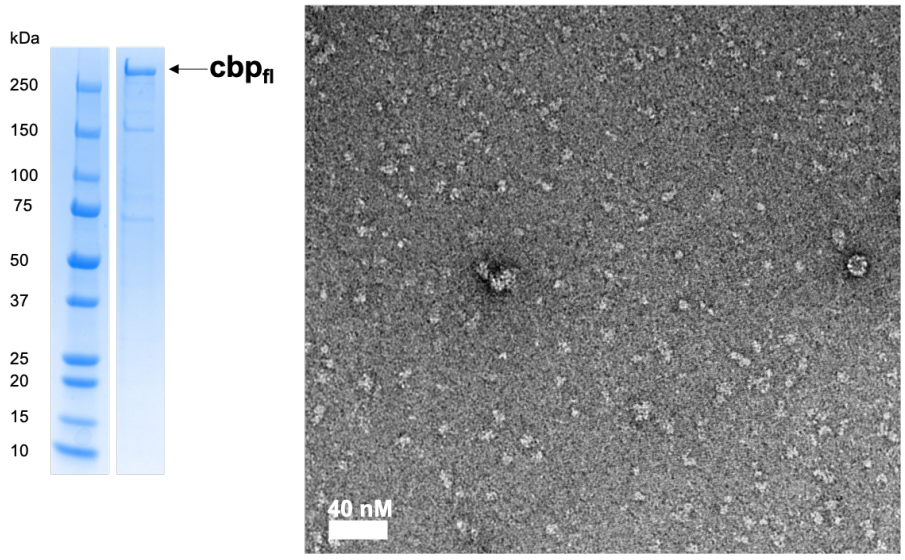


Figure 2.9. Purification of CBP_{fl} and **Negative Stain EM of Analysis of Purified CBP_{fl}**. SDS -PAGE analysis of purified CBPfl (left panel) and Representative Negative Stain EM Micrograph of the purified CBP_{fl} (right panel). 40nm scale bar shown on bottom left of micrograph.

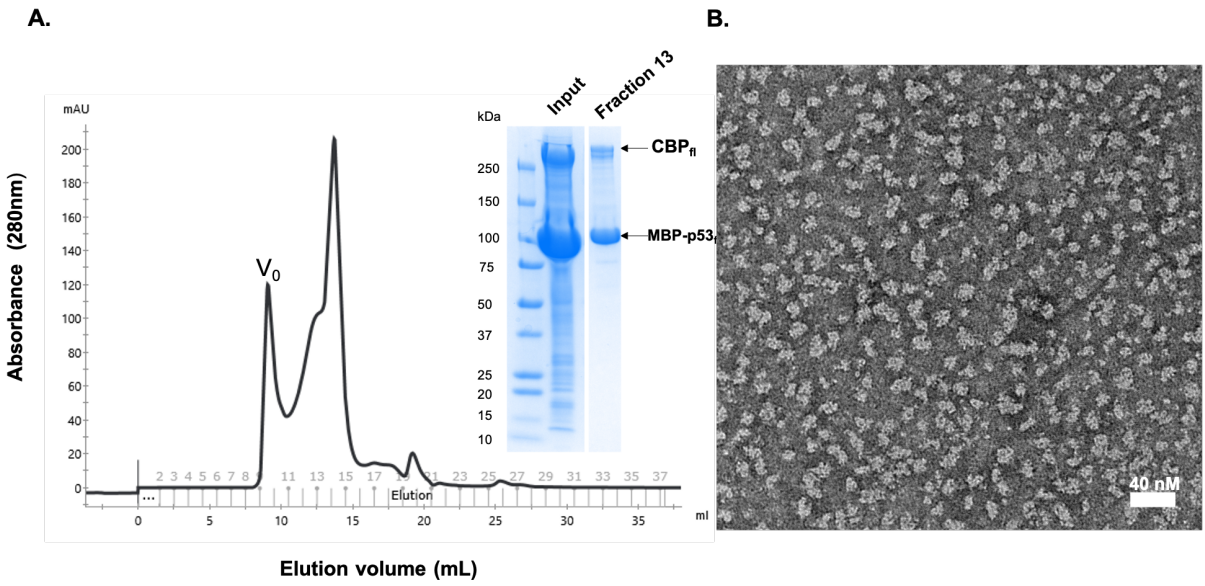


Figure 2.10. Purification of CBP•(p53)4 and Negative Stain EM of Analysis of Purified CBP•(p53)4. (A) Size exclusion chromatography profile of the CBP•(p53)4 binary complex run on a Superose 6 Increase 10/300 column and corresponding SDS -PAGE analysis of peak fraction. **(B)** Representative Negative Stain EM Micrograph of the purified CBP•(p53)4 binary complex**.** 40 nm scale bar shown on bottom right of micrograph.

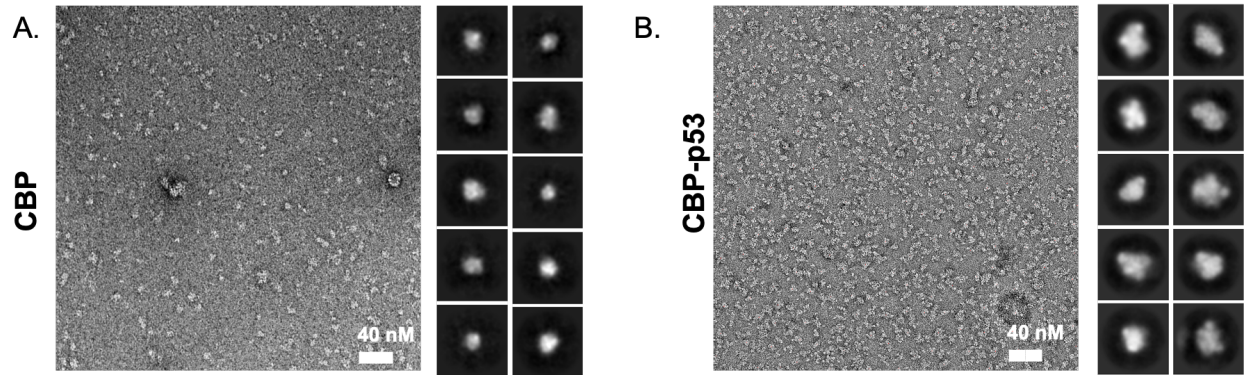


Figure 2.11 Single Particle Negative Stain EM Analysis and 2D Classification of CBP_{fl} and the $CBP•(p53)_4$complex. (A) Representative negative stain EM micrograph and 2D class averages of CBP_{fl} **(B)** Representative negative stain EM micrograph and 2D class averages of the $CBP•(p53)_4$complex**.** 40nm scale bar shown on bottom right of EM micrographs.

A Higher Calling: Stepwise Assembly and Characterization of the CBP-p53 DNA Coactivator Activator ternary complex by Negative Stain Electron Microscopy

I next analyzed EM images of the cbp-p300 binary complex reconstituted with DNA to determine the influence of DNA binding by activators on coactivator interactions. We hypothesized that the addition of DNA to the complex would further stabilize the binary complex, resulting in a less extended structure and further ordering of disordered regions within p53. For these experiments I selected 24-mer p53 consensus sites from either the p21 or gadd45 genes as the DNA substrate. Both p21 and gadd45 have been shown to be important p53 target genes and bind the DNA binding domain of p53 with K_d values in the low nanomolar range. Negative stain 2D averages from both preparations of the $CBP•(p53)_4$ binary complex with each DNA substrate revealed an architecture with striking resemblance to one another. Upon initial inspection, it appears that in both complexes with DNA, the overall shape of the complex becomes less extended and more compact, likely due to more ordering of the flexible portions of p53 **(Figure 2.12).** This conclusion is further supported by size exclusion profiles of the samples that when overlayed show slight shift in the elution positions indicative of differing hydrodynamic radii of the complexes **(Figure 2.13).** These findings are also consistent with previously published reports of p53 showing that in the free form, p53 adopts a highly dynamic open, cross shaped conformation. Upon DNA binding, this open conformation closes around the DNA double helix forming a compact complex. In both the free and DNA-bound p53

forms, the N termini remain extended with the proline-rich regions linking the TADs to the core domain projecting the TAD domains away from the central core of p53. As a result, the TAD domains are positioned in such a way to be freely targeted by a myriad of signaling proteins including CBP and p300 and to be subjected to extensive posttranslational modifications.

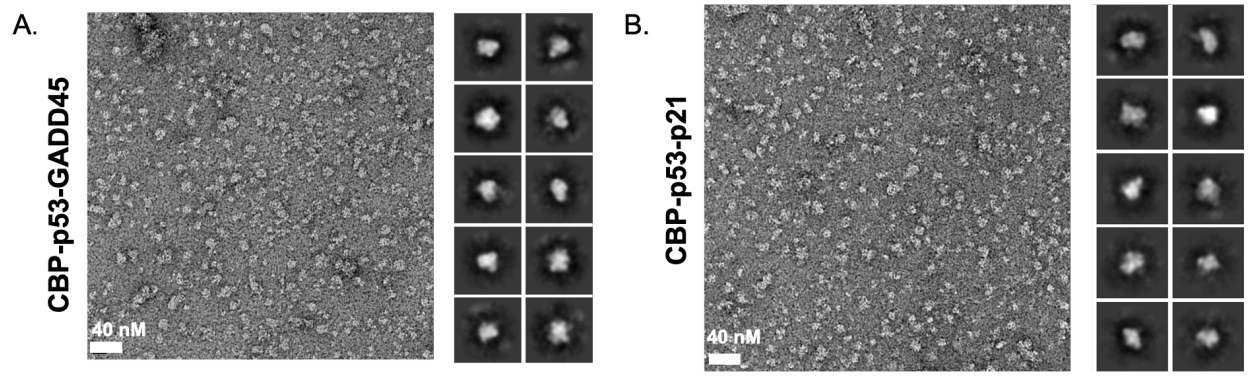


Figure 2.12 Single Particle Negative Stain EM Analysis and 2D Classification of CBP•$(p53)_4$-DNA complexes. **(A)**Representative negative stain EM micrograph and 2D class averages of the CBP•$(p53)_4$•gadd45 complex **(B)** Representative negative stain EM micrograph and 2D class averages of the CBP•$(p53)_4$•p21 complex**.** 40nm scale bar shown on bottom left of EM micrographs.

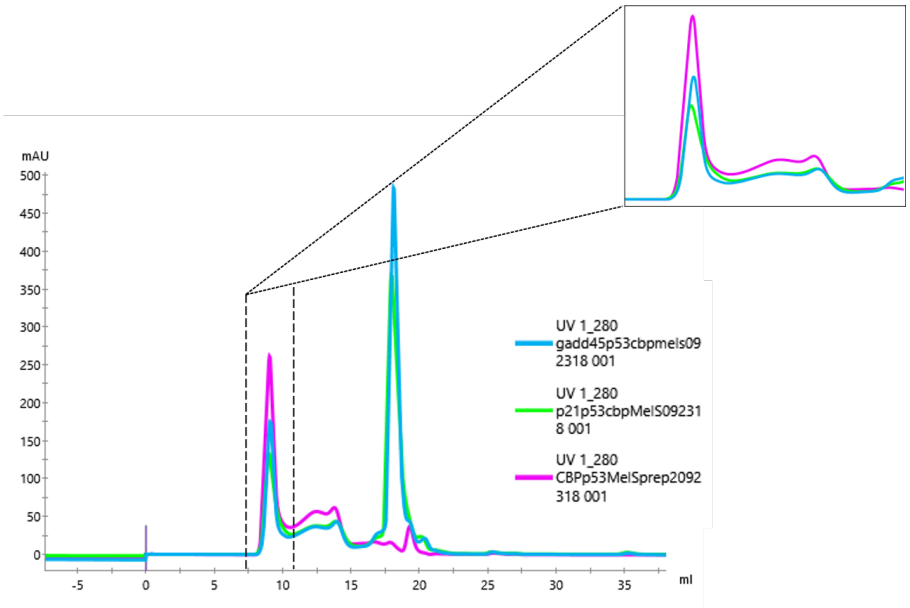


Figure 2.13. Purification and Analysis of CBP•$(p53)_4$•DNA Complexes. Size exclusion chromatography showing the overlayed elution profiles of the CBP•$(p53)_4$, binary complex (pink), CBP•$(p53)_4$•gadd45, (blue) and the CBP•$(p53)_4$•p21 (green) run on a Superose 6 Increase 10/300 column.

2.3 Conclusions

All They See is Rodneys. All I see is KINGS

Here I present the biochemical and structural characterization of an intact transcriptional activator coactivator complex formed between CBP and the tumor suppressor p53 reconstituted from recombinantly expressed and purified proteins. To obtain sample for biochemical characterization, I expressed the human CBP•$(p53)_4$ binary complex through baculovirus-mediated coinfection and overexpression in insect cells and demonstrated. From these preparations, I was able to isolate and purify the complex from cells to near homogeneity and demonstrate in vitro for the first time a 4:1 p53:CBP interaction. Furthermore, I have been able to show using negative stain microscopy that CBP undergoes structural rearrangement upon complexation with p53 and this complex is further modulated the incorporation of the cognate DNA sequences for p53. The EM analysis provides a critical snapshot of the intact CBP•$(p53)_4$ activator coactivator complex and evidence for a highly dynamic complex with conformational and structural heterogeneity. While these studies have provided an initial framework for understanding the overall assembly and the dynamic nature this system, the limited resolution of our initial 2D classifications does not allow clear structural features to be observed and one cannot describe the interactions present in significant detail. In spite of this, the insights gained here will not only guide, but will be key, for the successful characterization CBP and p300 at high-resolution by cryo-EM. Furthermore, the approaches utilized will prove useful for studying other signaling complexes involving the coactivator CBP and its cognate binding partners at large.

2.4 Experimental Methods

Expression Trials in Insect Cells.

Constructs for full-length human CBP, p300 and p53, were obtained and each construct was transformed into Max Efficiency DH10Bac competent *E. coli* to generate recombinant bacmids (baculovirus shuttle vectors) and ultimately baculoviral stocks for protein expression. Optimal expression conditions were conducted by setting up small-scale insect expression cultures (5mL) in 24 well deep blocks. The cells were incubated at 20°C or 27°C with the following conditions varied: media additives, multiplicity of infection (MOI) or incubation times (24, 48, 72, and 92 hour). Small scale batch nickel purifications were used to capture the proteins. Protein expression was assessed by visualization of Coomassie Blue Stained SDS-PAGE.

Protein Purification (Final Biochemical Conditions).

N terminally His6-FLAG tagged CBP and MBP-tagged-p53 baculovirus were individually prepared and amplified by using the Bac-to-Bac baculovirus expression system. Sf9 or Hi Five cells at a cell density of 2 *10^6 insect were co-transfected with the appropriate virus and incubated with shaking (120 rpm) at 27 °C for 24 hours (Hi Five cells) or 48 hours (SF9 cells) before collection. For purification cells were lysed in 20 mM Tris-HCl pH 7.4, 300 mM NaCl, 35 mM imidazole, 2.0 mM $MgCl_2$, and 0.5 mM PMSF. The lysis buffer was supplemented with 1 protease EDTA-free inhibitor tablet, 5 u/mL benzonase, and 250 µL/mL of insect cell pop culture reagent (for chemical lysis). The supernatant was clarified with centrifugation and loading onto a 5 ml Ni-Affinity HisTrap HP column (GE healthcare) using an Akta purifier FPLC. The column was washed with Nickel buffer A (20mM Tris-HCL pH 7.4, 300mM NaCl, 35mM imidazole, 2mM TCEP) and eluted with Nickel Buffer B (20mM Tris-HCL pH 7.4, 300mM NaCl, 500 mM imidazole, 2mM TECP). Fractions corresponding to eluted protein were collected and loaded onto a gravity flow column packed with 3 mL of Flag-M2 agarose (Sigma) and equilibrated with Flag Column Buffer (20mM Tris-HCL pH 7.4, 300mM NaCl, and 2mM TCEP). The flow through was collected and re-loaded onto the packed column two more times to ensure maximum yield of purified complex. The loaded column was washed three times with buffer A, and bound protein was eluted with 3x Flag elution buffer (Flag Column Buffer + 100ug/ml 3x Flag-peptide (Sigma). Fractions were pooled and concentrated using a Vivaspin centrifugal

concentrator (GE Healthcare) and loaded on a Superose 6 Increase 10/300 GL size exclusion column pre-equilibrated (20 mM tris pH 7.4, 300 mM NaCl, 50 uM $ZnCl_2$,1% glycerol, and 1mM TCEP. All proteins were analyzed by SDS-PAGE stained with Coomassie brilliant blue and visualized with negative stain EM.

Specimen Preparation and EM imaging on Negative Stained Samples.

3.5 µL of sample was applied to a glow-discharged continuous carbon grid. After a 1-minute adsorption the sample was blotted with filter paper, negatively stained with three consecutive droplets of 2% (w/v) uranyl formate solution, blotted again to remove residual stain, and air-dried. Filter paper was used to blot the grid. Specimens were imaged at room temperature with a Tecnai T12 electron microscope operated at 120kV. Images were recorded at magnifications ranging from 29,000X-62,000X.

*Two-Dimensional Classification***.**

Individual particles were manually picked and windowed using 324-pixel boxes. The particle projections were then subject to two-dimensional reference free alignment and classification in cryoSPACv215. Data sets include CBP (1,835 particles), CBP-p53 (3,156 particles), CBP-p53-gadd45 (4,423) and CBP-p53-p21 (2,894 particles).

2.5 References

1. Avantaggiati, M.L., et al., Recruitment of p300/CBP in p53-Dependent Signal Pathways. Cell, 1997. 89(7): p. 1175-1184.
2. Bekesi, A., et al., Challenges in the Structural-Functional Characterization of Multidomain, Partially Disordered Proteins CBP and p300: Preparing Native Proteins and Developing Nanobody Tools. Methods Enzymol, 2018. 611: p. 607-675.
3. Bose, D.A., et al., RNA Binding to CBP Stimulates Histone Acetylation and Transcription. Cell, 2017. 168(1-2): p. 135-149.e22.
4. Grossman, S.R., p300/CBP/p53 interaction and regulation of the p53 response. Eur J Biochem, 2001. 268(10): p. 2773-8.
5. King, L.A., R. Hitchman, and R.D. Possee, Recombinant Baculovirus Isolation, in Methods in Molecular Biology. 2016, Springer New York. p. 73-94.
6. Lee, C.W., et al., Mapping the Interactions of the p53 Transactivation Domain with the KIX Domain of CBP†. Biochemistry, 2009. 48(10): p. 2115-2124.
7. Lee, C.W., et al., Structure of the p53 Transactivation Domain in Complex with the Nuclear Receptor Coactivator Binding Domain of CREB Binding Protein. Biochemistry, 2010. 49(46): p. 9964-9971.
8. Miller Jenkins, L.M., et al., Characterization of the p300 Taz2–p53 TAD2 Complex and Comparison with the p300 Taz2–p53 TAD1 Complex. Biochemistry, 2015. 54(11): p. 2001-2010.
9. Ohi, M., et al., Negative staining and image classification — powerful tools in modern electron microscopy. Biological Procedures Online, 2004. 6(1): p. 23-34.
10. Peisley, A. and G. Skiniotis, 2D Projection Analysis of GPCR Complexes by Negative Stain Electron Microscopy, in Methods in Molecular Biology. 2015, Springer New York. p. 29-38.
11. Punjani, A., et al., cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nature Methods, 2017. 14(3): p. 290-296.
12. Skiniotis, G. and D.R. Southworth, Single-particle cryo-electron microscopy of macromolecular complexes. Microscopy (Oxf), 2016. 65(1): p. 9-22.
13. Stark, H. and A. Chari, Sample preparation of biological macromolecular assemblies for the determination of high-resolution structures by cryo-electron microscopy. Microscopy, 2016. 65(1): p. 23-34.
14. Stolt-Bergner, P., et al., Baculovirus-driven protein expression in insect cells: A benchmarking study. Journal of Structural Biology, 2018. 203(2): p. 71-80.
15. Takizawa, Y., et al., While the revolution will not be crystallized, biochemistry reigns supreme. Protein Science, 2017. 26(1): p. 69-81.
16. Teufel, D.P., et al., Four domains of p300 each bind tightly to a sequence spanning both transactivation subdomains of p53. Proceedings of the National Academy of Sciences, 2007. 104(17): p. 7009-7014.
17. Yu, Q., et al., Global Conformational Selection and Local Induced Fit for the Recognition between Intrinsic Disordered p53 and CBP. PLoS ONE, 2013. 8(3): p. e59627

CHAPTER THREE

Making Meaning, A Critical and Thoughtful Examination of the Structures of Life: Structure Determination and Analysis of the Human CBP-p53 Complex

Abstract

The Revolutionist: Born a Slave. Crowned a King. Currently Living Life Legendary.

Gene transcription is precisely orchestrated from the outputs of an intricately woven dynamic protein-protein interaction (PPI) network. At the heart of this PPI transcriptional regulatory network are interactions formed between coactivators like CBP and gene-specific transcriptional factors such as p53. Despite being central coordinators of transcription, our understanding of the mechanisms of activation remain restricted, due to a lack of structural information on full-length coactivator complexes. Here, I utilized single particle cryo-EM to determine and visualize the first three-dimensional structure of the full-length human CBP-p53 coactivator-activator complex. This analysis not only captures the overall architecture of but reveals that the CBP-p53 binary complex exists in a series of distinct structural states. Collectively, the observations from this study provide new and exciting insights into the structural arrangement and underlying conformational dynamics of coactivator complexes, both of which are likely to have important functional implications for DNA binding and transcriptional regulation.

3.1 Introduction

Connecting a Vision of Tomorrow Through Theory and Practice Today
"I started from the Bottom"

The regulation of gene expression is a complex task that is critical for the growth development, and survival of all organisms. This endeavor is achieved through protein–protein interactions (PPIs) formed between transcriptional activators and coactivators. Coactivators form a special class of regulatory proteins capable of physically interacting with an abundance of other regulatory proteins including DNA-binding activators Mediator, RNAPII, and other molecular players of the general transcriptional machinery. In light of coactivators unique facility to bind to such an expansive catalogue of proteins, it has been long speculated this molecular aptitude is due in part to the presence of intrinsically disordered regions (IDRS). Practically speaking, IDRS confer coactivators with a degree of structural plasticity that permits structural modulation by their binding partners with relative ease. Adjustments of coactivator confirmation thus serves as one point of control over their activity. Similarly, it has been proposed that the complexation of a coactivator with a given activator culminates in a distinct series of substructures further enhancing the regulatory potential of the coactivator. Although it is tempting to envision a transcription scenario that fully embraces these conceptual models, our understanding of the mechanics of transcription will continue to remain incomplete and potentially inaccurate without detailed structural knowledge of the structures of activator bound coactivator assemblies. Unfortunately, coactivators and their complexes have proven to be challenging with respect to their structural characterization largely owing to their size, inherent structural flexibility, and instability. Our understanding of the structural basis of transcription activation by CBP and p300 provides an illustrative example of the promise and peril of the structural characterization of these integral regulators as highlighted below.

The master coactivators, CREB-binding protein (CBP) and its paralog p300 are critical coordinators of gene expression and provide an excellent example of the signature roles assumed by coactivators. CBP and p300 alter the chromatin landscape as readers and writers of epigenetic marks, bind to zillions of sequence specific transcription factors, and act as a bridge between the activators and the general transcription machinery.

Influencing nearly every stage of transcription, through these multifaceted roles, CBP and p300 integrate numerous signaling pathways to execute a unique transcriptional outcome tailored to serve the needs of the cell in a spatially and temporary regulated manner.

A growing body of biochemical studies and high-resolution structural data have emerged mapping domain wise interactions between CBP or p300 and activators. One well-studied interaction is with the tumor suppressor protein p53. While investigations on this functionally critical complex have provided crucial insights, lending support to the prevailing model of the field (as described in chapter 2), they do not capture the entire spectrum of interactions and molecular contacts contributing to the formation of the complex which will be required to definitely define a mechanistic model. Neither do the studies provide full models of the conformational dynamics underpinning complex assembly. In particular, few investigations have attempted to resolve unambiguously outstanding questions regarding the possibility of multiple structural scaffolds arising from the incorporation of multiple and divergent activators binding to the ABDS of one or more molecules of CBP or p300 with distinct stoichiometries. Consequentially, our understanding of the mechanisms underlying p53 transcriptional activation by CBP or p300 still remain obscure due to the absence of a complete structural depiction of the complex. Therefore, the structural characterization of an intact activator coactivator complexes represents a much needed, yet missing link to fully define the structural and mechanistic basic of coactivator function and how it changes with activator binding.

Towards this goal, I successfully isolated and determined the overall cryo-EM structure of the full-length human CBP-p53 coactivator-activator complex. Structural analysis reveals two predominant conformational arrangements of the conformers at moderate estimated resolutions of 7.2 Å and 8.2 Å for the first conformer and 7.2 Å and 10.9 Å for the second conformer. Taken together, the architectural organization and visualization of conformation dynamics for this complex provide direct evidence for a mechanism whereby structural transitions within coactivators and the resulting complexes they form with activators serve as a key point of control to fine-tune gene expression profiles in response to developmental and environmental cues.

3.2 Results and Discussion

From the Ground Up: Isolation of the human CBP-p53 binary complex and Cryo EM "And now I'm Here"

The proteolytic sensitivity and flexible nature of CBP and p300 has long hampered isolation and moderate to high resolution structure determination of these critical commanders of transcription. Through co-expression of full-length human CBP and full-length p53 in Spordoptera frugiperda (SF9) cells (or in some studies Hi-Five Cells) I was able to purify the intact CBP-p53 binary complex with a total approximate molecular weight of 615 kDa (tagged versions of the complex). As outlined in Chapter Two a fine-tuned baculovirus infection and optimized purification strategy enabled me to obtain a relatively homogenous biochemical sample. The final purification scheme **(Figure 3.1 A),** for cryo-EM investigations consisted of lysis and isolation of the CBP-p53 binary complex from SF9 cells, followed by immobilized metal affinity chromatography, a flag affinity purification step, and a final size exclusion chromatography step to polish purity of the resulting sample and to assess complex formation. SDS-PAGE analysis of the peak fraction corresponding to the approximate molecular weight of a complex with 4:1 p53:CBP ratio revealed a sample that was purified to near homogeneity and that appeared stoichiometrically accurate **(Figure 3.1 B-C)**. Given the biochemical evidence that the human CBP-p53 complex co-expressed from insect cells formed a homogenous and stable assembly, I next examined this sample using single particle EM. As a first step, the complex was screened by negative stain EM to further assess homogeneity and oligomeric state. Inspection of the negative stain EM micrographs revealed a relatively homogenous field of particles with a size and shape consistent with that of an intact CBP-p53 coactivator complex **(Figure 3.1D)** suitable for cryo-EM studies. Unlike the previous data set collected on this complex with negative stain EM, cryo-EM permits imaging of samples in their fully hydrated states, providing the potential for higher resolution structural information. The technique also holds potential to uncover the structural dynamics of macromolecular complexes through the implementation of powerful 3D classification. Accordingly, I employed single particle cryo-EM in order to gain a more in-depth insight into the overall architecture and conformational rearrangements of the human CBP-p53 core complex. It is important to note that the success of the cryo-EM

was not only critically dependent on obtaining CBP-p53 binary complex preparations that were as homogenous as possible but was highly contingent upon time. The sensitive nature of the complex limited the timeframe that the sample could be confidently used in experiments. Thus, all three purification stages were completed on a single day to limit degradation of purified complex. Furthermore cryo-EM grid preparation was done immediately after the sample purification and screening.

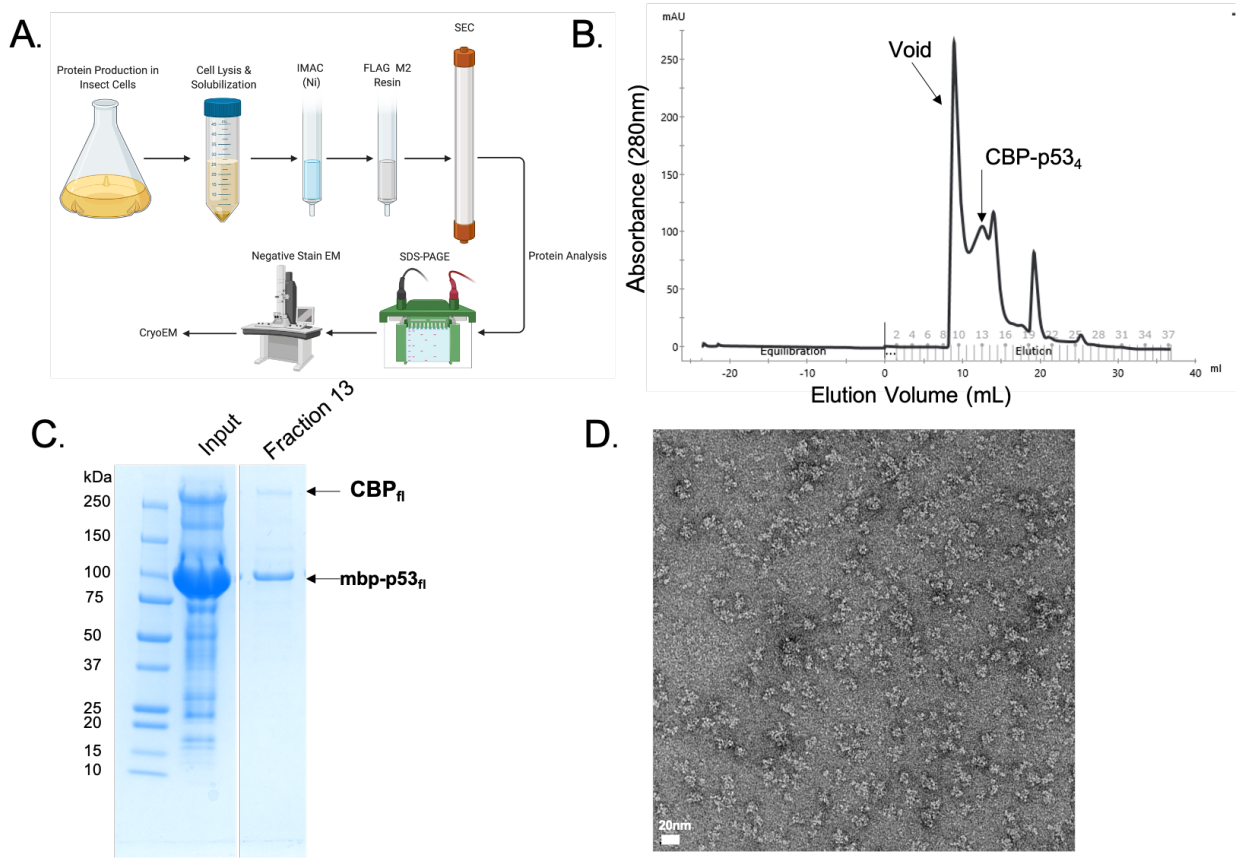


Figure 3.1 Biochemical Reconstitution of the Human CBP-p53 Transcriptional Coactivator-Activator Complex. (A) Expression and Purification Process. Figure created with BioRender.com (B) Size-exclusion chromatography of recombinant CBP-p53 complex and (C) Coomassie Brilliant Blue SDS-PAGE. The molecular mass of is 605 kDa (D) Negative-stain EM of the recombinant cbp-p53 complex. Individual particles have triangular shape with a length of ~18 nm and a thickness of ~12nm. 20nm scale bar shown at bottom left corner of micrograph.

For cryo-EM, several thousand micrographs were collected. Movies were subjected to beam-induced motion correction, and the contrast transfer function parameters for each micrograph was estimated. From this, 2,412 micrographs were selected for further processing based on their overall quality after visual inspection of CTF fit and average defocus values (**Table 3.1**) Initial reference-free three class averages

were carried out both in cryoSPARC and RELION data processing software packages. No apparent differences in quality or appearance could be detected between the independently determined particle projections **(Figure 3.2)** Therefore cryoSPARC was arbitrarily chosen to carry out all data processing steps**.**

Table 3.1 Data Collection and Processing

Structure: Human CBP-p53		
Data Collection		
Microscope	Talos Artica	
Magnification	45,000	
Voltage (kV)	200	
Stage Tilt (°)	0	
Detector	K2 Summit	
Recording Mode	Counting	
Total Electron Exposure (e-/Å^2)	64.15	
Number of Frames	40	
Defocus Range (uM)	-1.5-3.0	
Pixel Size (Å)	0.91	
Data Processing		
	Conformation One	**Conformation Two**
Data Processing Platform	cryoSPARC v2.15	cryoSPARC v2.15
Number of Micrographs	2,412	2,412
Final Particle Images (no.)	49,548	46,255
Symmetry Imposed	C1	C1
Refinement Map Resolution (Å)	8.2	11.0
Map Resolution Range (Å)	4.2-16.9	15.4-7.8
Local Refinement Map Resolution (Å)	7.2	7.2
Local Map Refinement Resolution Range	6.6-15.8	9.9-15.4
FSC threshold	0.143	0.143

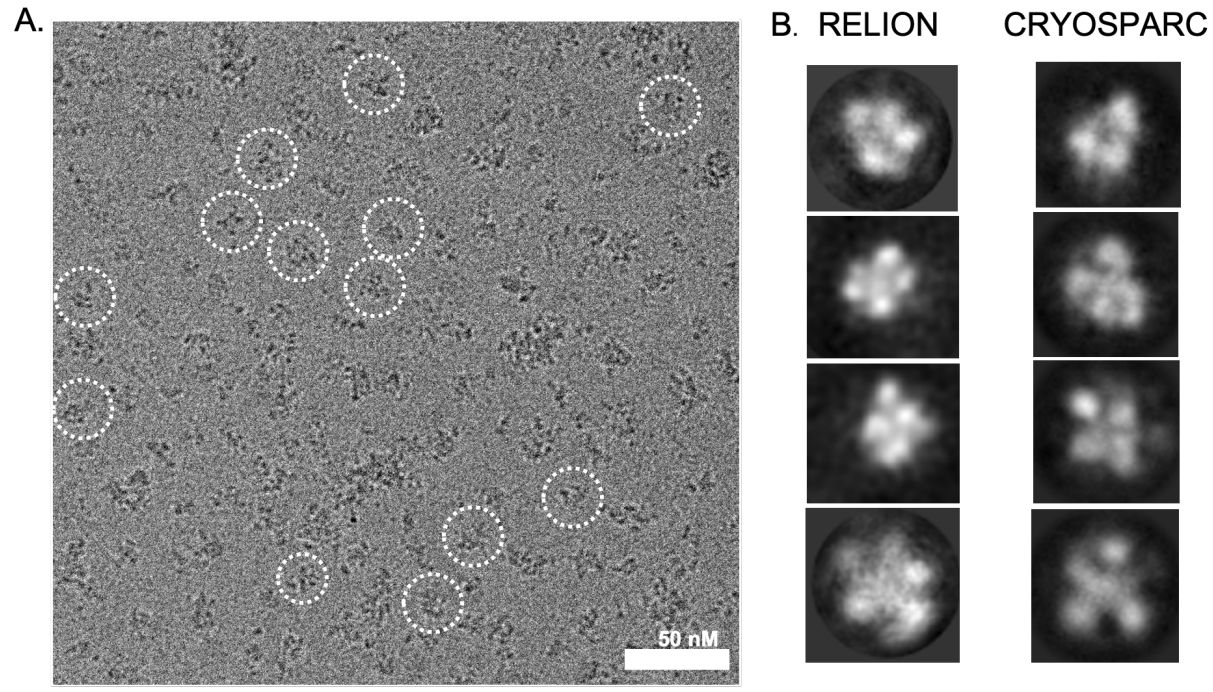


Figure 3.2 Cryo-EM Micrograph and Two-Dimensional Class Averages of the human CBP•$(p53)_4$ complex. (A) Representative cryo-EM micrograph showing the human CBP•$(p53)_4$ trapped ice. Example of particle projections are circled in white**.** Scale bar, 50nM**. (B)** Representative initial reference-free 2D class averages of the human CBP•$(p53)_4$ complex processed in RELION (left panel) and cryoSPARC (right panel).

Mapping a Path to Success: Architecture of the human CBP-p53 Transcriptional Activator Coactivator Complex

Previous studies have suggested that conformational changes are an essential component of CBP/p300 and p53 mechanism. Based on this premise, I hypothesized that the human CBP-p53 complex would likely exists in a number of distinct structural states. Unsurprisingly, inspection of the two-dimensional (2D) image analysis and classification of the CBP-p53 complex reveal flexible conformers which would be consistent with this notion **(Figure 3.3)**. For this reason, cryo-EM reconstructions for this complex involved a multi-model three-dimensional (3D) refinement strategy (**Figure 3.4- Figure 3.5**) Because of extensive 2D projection and particle sorting early on, the cryo EM data set could be subdivided into two particle stacks acquired from optimized automated picking parameters. This helped to further facilitate the partitioning of the EM data set into distinct structures. Each distinct particle stack was then subjected to reference-free *ab initio* reconstructions. Initial models from ab initio reconstructions were then selected on the basis of clear features present enabling a multimodal refinement strategy for cryo-EM

reconstruction of these samples. The partitioned particle sets were processed and refined separately ultimately producing two clearly structurally unique conformations of the CBP-p53 complex. (**Figure 3.4-Figure 3.5**) An asymmetric quadripartite configuration was evident in both resulting models. The resolution of each structure was determined by the Gold Standard Fourier Shell Correlation (FSC) with 0.143 cutoff criteria. It is important to make note of the sharp drop observed in the FSC curves is likely due to challenges with alignment due to the disorder present disorder present, and overall structural heterogeneity of this biological. Further validation on the structural models presented will be required in order to ensure the accuracy of the structural models presented.

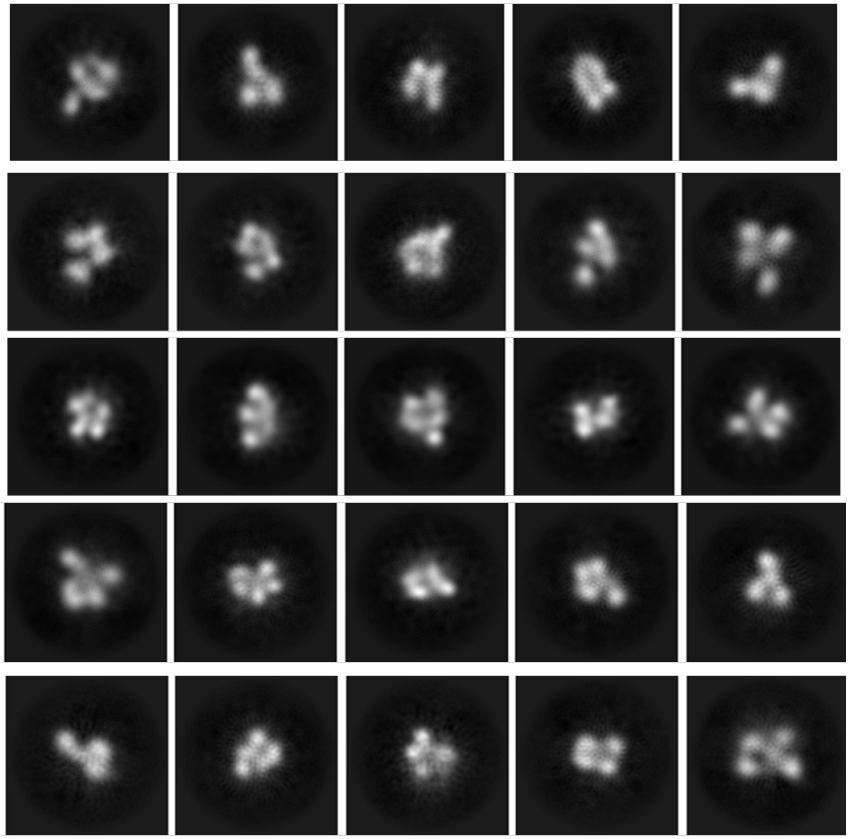


Figure 3.3. Reference-free Two-Dimensional (2D) Class Averages Generated from the cryo-EM data set on the human CBP•$(p53)_4$ complex. Representative reference free 2D classes show distinct shapes and arrangements**.**

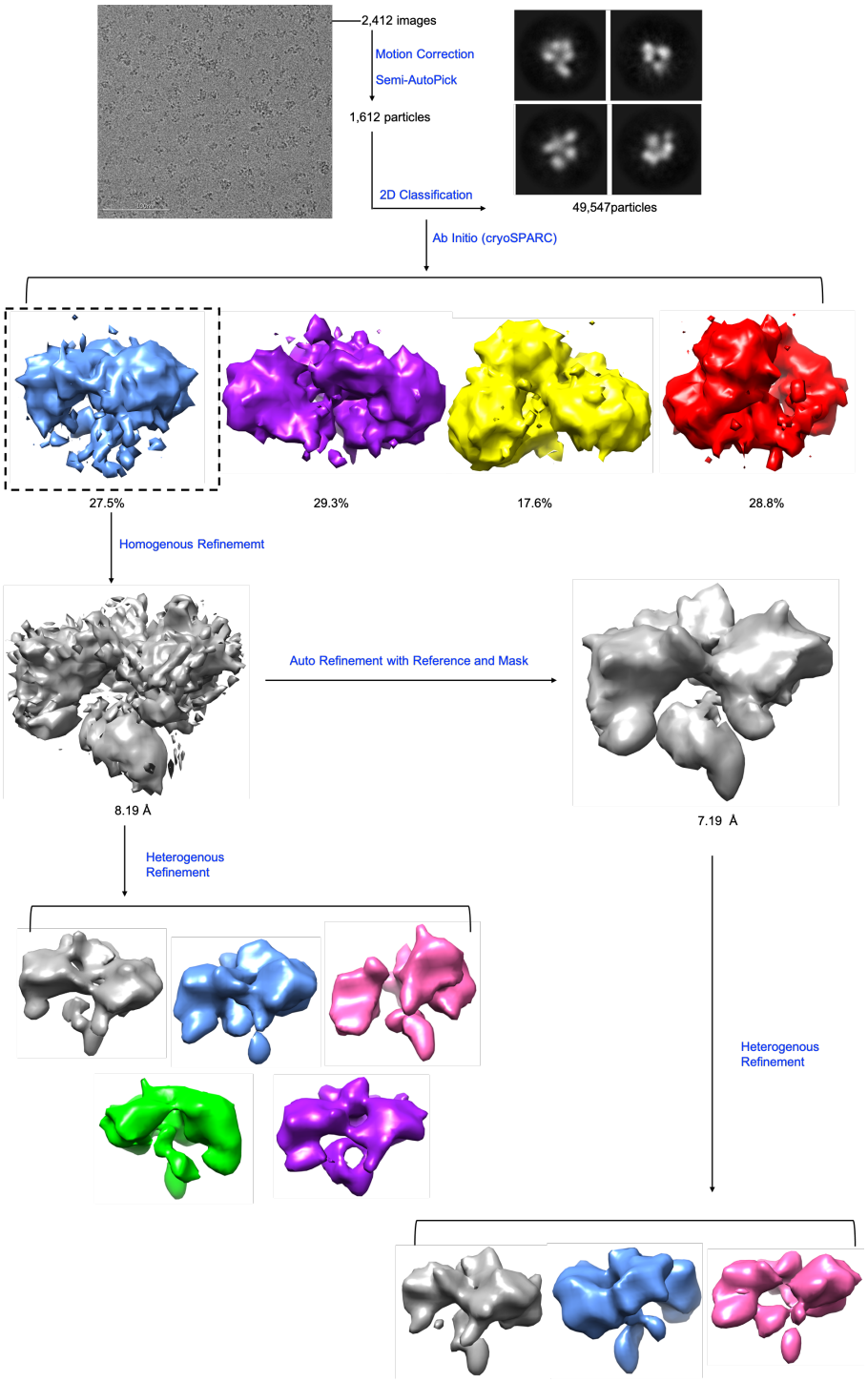


Figure 3.5 Data Processing Strategy and Pipeline for Single Particle Cryo-EM Analysis of the human CBP•$(p53)_4$ (conformation one).

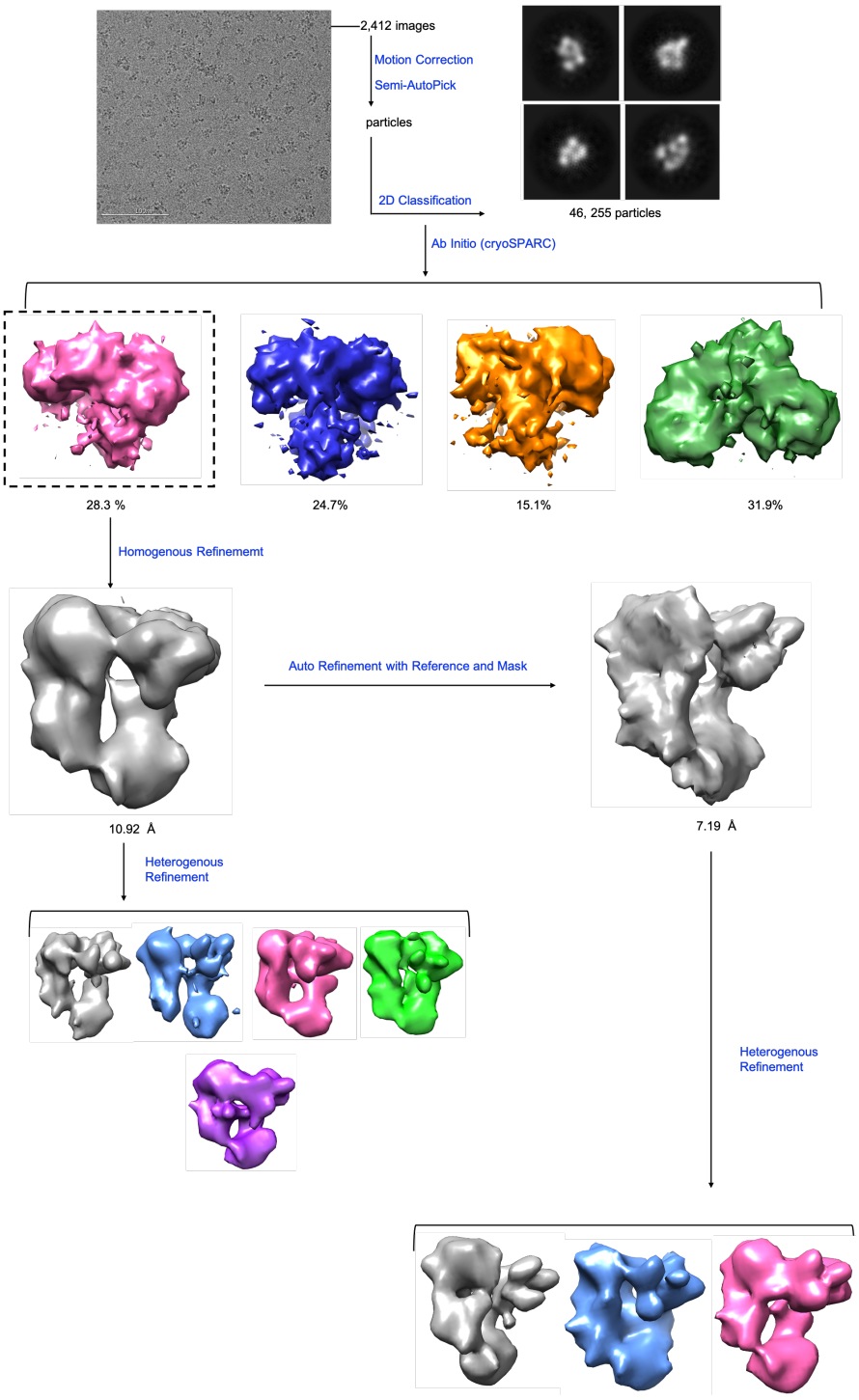


Figure 3.6 Data Processing Strategy and Pipeline for Single Particle Cryo-EM Analysis of the human CBP•$(p53)_4$ (conformation two).

With no symmetry-imposed refinement of the first conformer enabled a three-dimensional (3D) reconstruction with overall global 8.2 Å resolution. Through a secondary automated refinement inclusive of the mask output from the first refinement an overall global resolution of 7.2 Å for this same reconstruction was acquired **(Figures 3.6-3.8).**

As shown in **Figure 3.9**, the first configuration of the human CBP-p53 has the appearance of a three-leaf shamrock. A quasi-symmetrical density rich upper region forms the flowering portion of the shamrock with one large and two smaller leaf-like lobes (numbered clockwise from bottom left of right when viewed from the top **(Figures 3.10-3.11**). Each lobe of the density rich flowering portion are connected via narrow junctions forming this. The second of the smaller lobes connects loosely to the stem through a small linker. Minor crevices and protrusions are present throughout the topology of the structure.

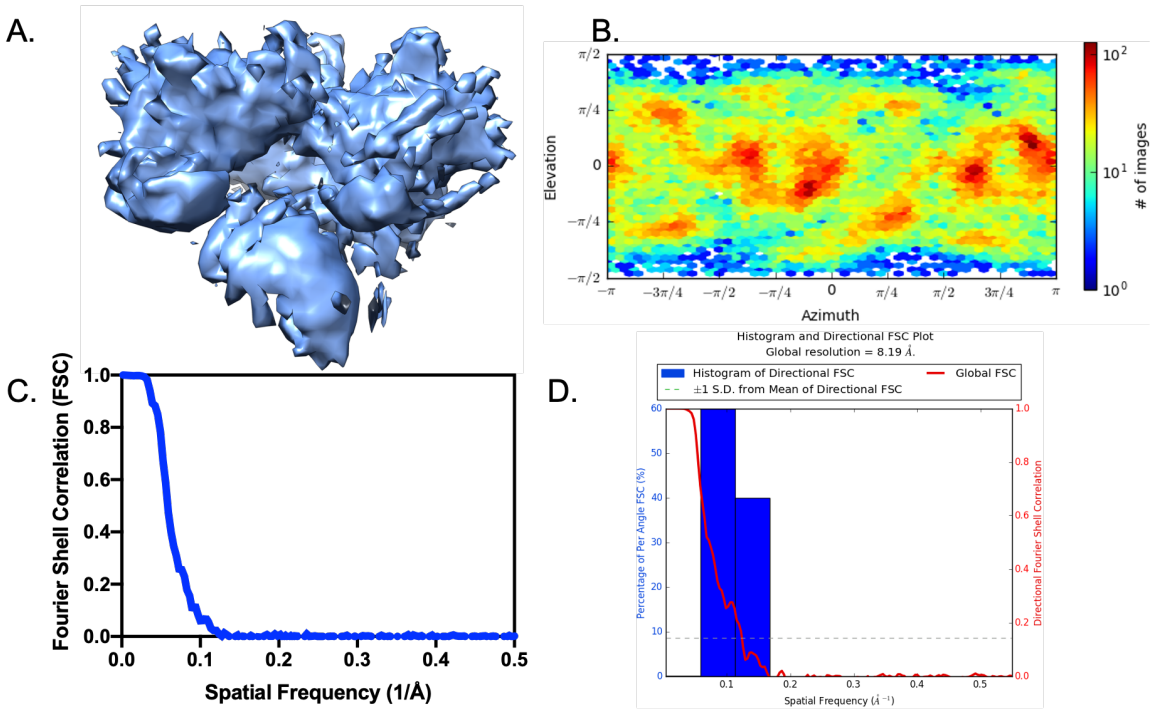


Figure 3.6 Cryo-EM Three-Dimensional (3D) Reconstruction of the Human CBP•$(p53)_4$ Binary Complex (Conformation One) and Validation of cryo-EM Reconstruction. (A) Three-dimensional reconstruction of the human CBP•$(p53)_4$ (conformation one) binary complex. The estimated resolution of the map is 8.2 Å. **(B)** Euler angle distribution of the human CBP•$(p53)_4$ (conformation one) binary complex highlighting some preferred orientations. **(C)** Gold Standard FSC curve for conformation one of the human CBP•$(p53)_4$. **(D)** Directional FSC histogram for the final 3D reconstruction of the complex.

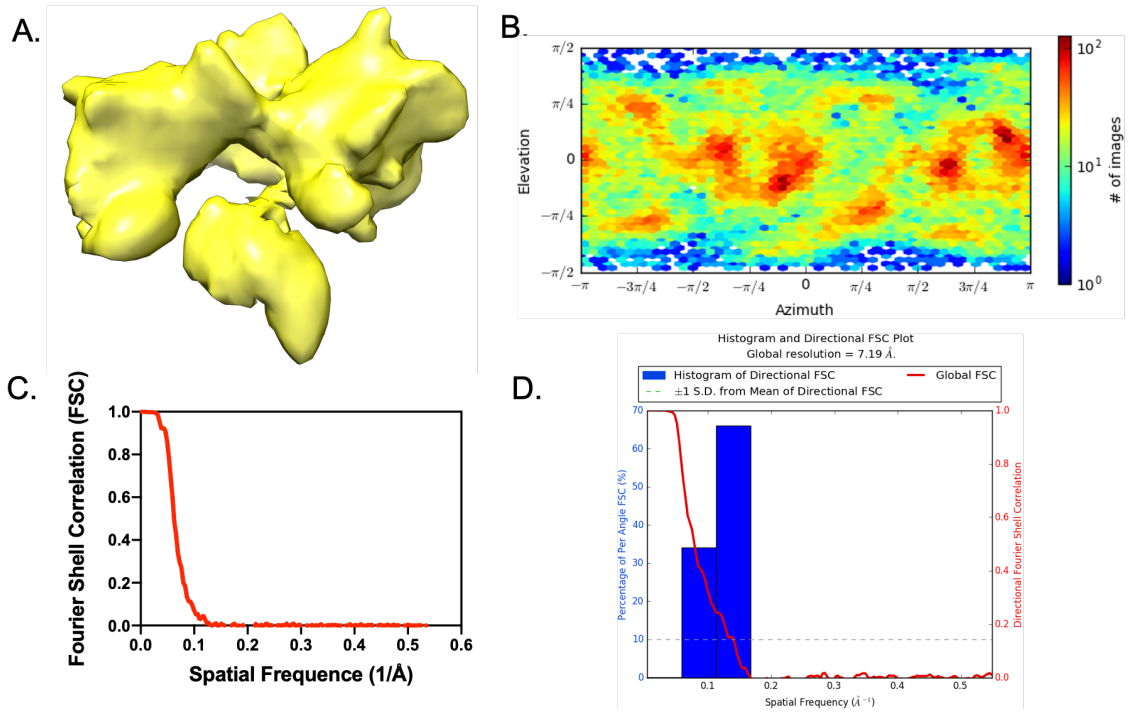


Figure 3.7 Cryo-EM Three-Dimensional Reconstruction of the Human CBP•$(p53)_4$ Binary Complex (Conformation One) and Validation of cryo-EM Reconstruction After Further Refinement. (A) Three-dimensional reconstruction of the human CBP•$(p53)_4$ (conformation one) binary complex. The estimated resolution of the map is 7.2 Å. **(B)** Euler angle distribution of the human CBP•$(p53)_4$ (conformation one) binary complex highlighting some preferred orientations of the compex. **(C)** Gold Standard Fourier Shell Correlation (FSC) curve for conformation one of the human CBP•$(p53)_4$. The final resolution was determined using FSC criterion FSC=0.143 criterion. **(D)** Directional FSC histogram for the final 3D reconstruction of the complex.

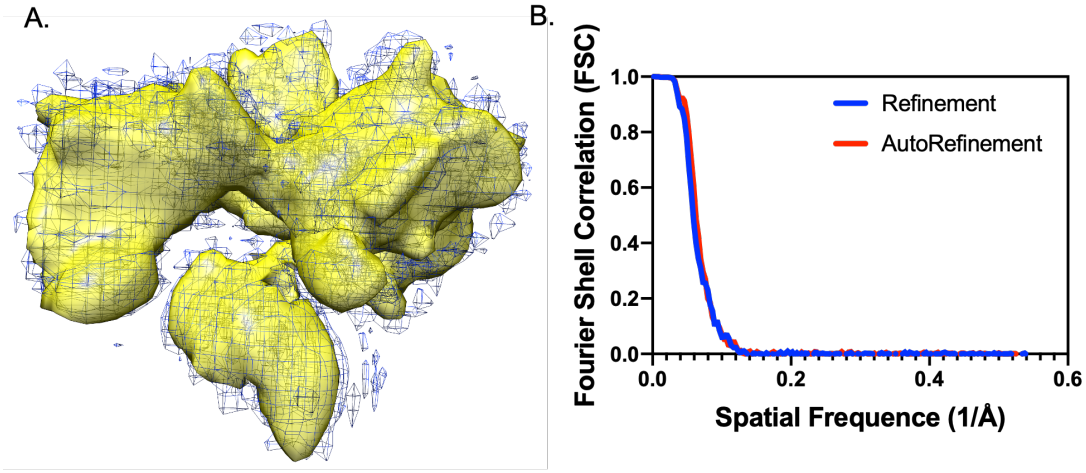


Figure 3.8 Human CBP•$(p53)_4$ Binary Complex (Conformation One) Refinement Comparisons.
(A) The cryo-EM density map for the initially refined structure of the human CBP•$(p53)_4$ binary complex (blue mesh is) superimposed on the final structure refinement (yellow solid). **(B)** Overlayed initial refinement (blue) and final refinement **(red)** Gold Standard Fourier Shell Correlation (FSC) curves for conformation one of the human CBP•$(p53)_4$ complex.

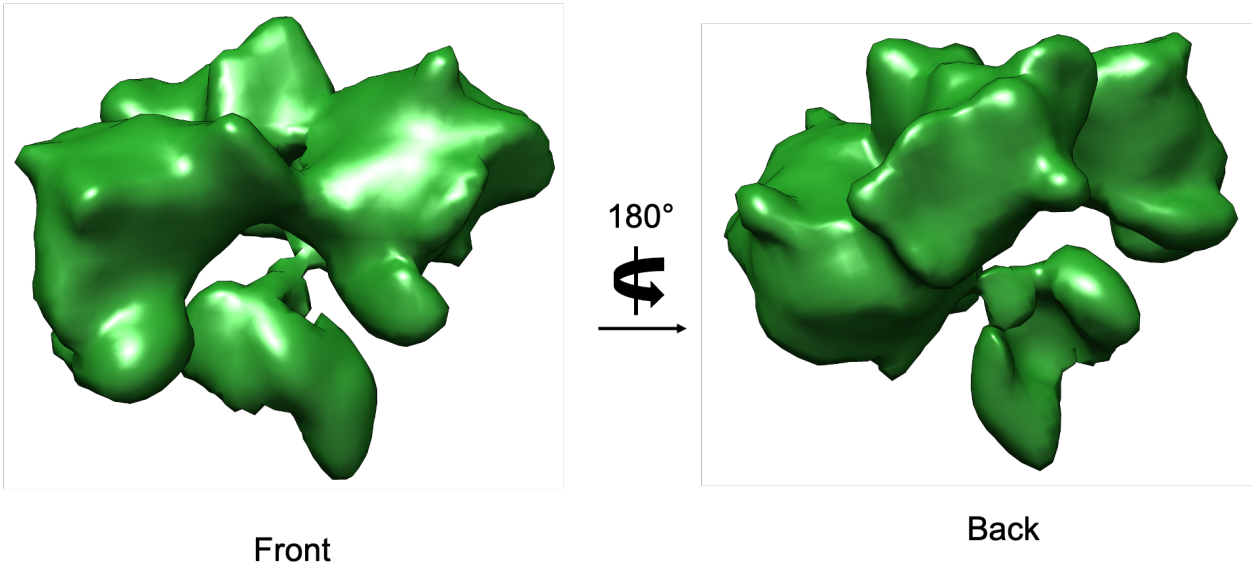


Figure 3.9. Overall Structure of The Human CBP•$(p53)_4$ Binary Complex (Conformation One). The human CBP•$(p53)_4$ binary complex (conformation one) adopts a quadripartite structure that resembles a three-leaf shamrock.

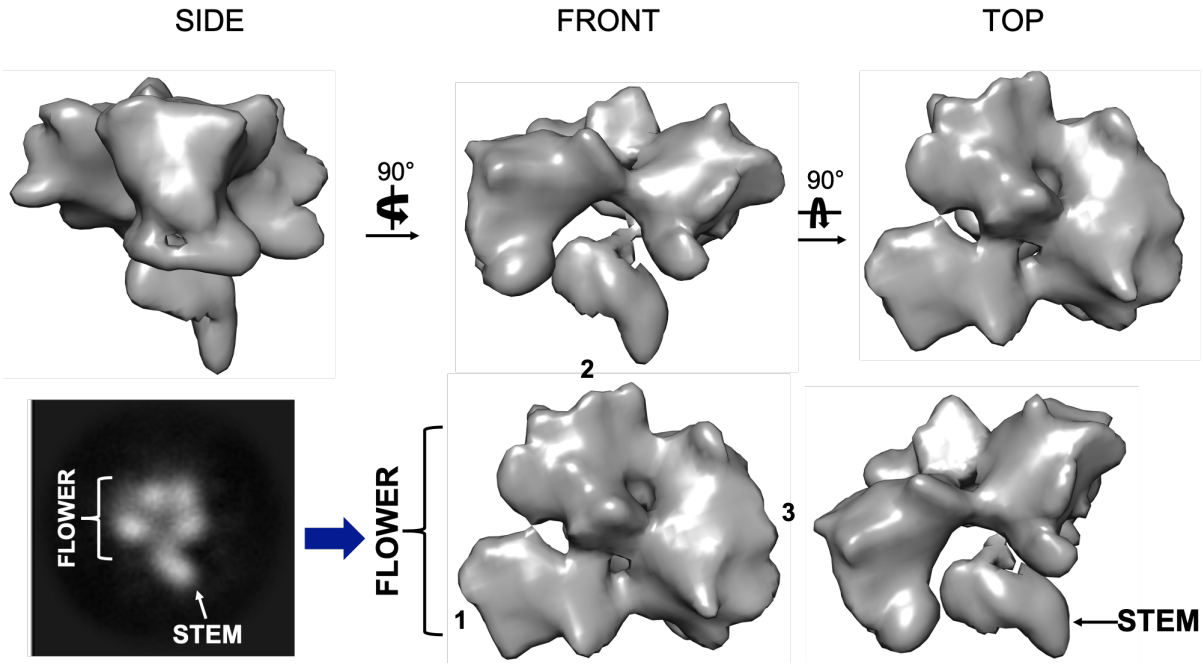


Figure 3.10. Structural Features of the Human CBP•$(p53)_4$ Binary Complex (Conformation One). The overall architecture of the human CBP•$(p53)_4$ binary complex (conformation one) resembles a three-leaf shamrock (flower), consisting of an upper region with three density rich lobes and a smaller bottom lobe (stem). Depiction of the 3D reconstruction is shown from three viewing angles side, front, and top (upper panel) and is compared with a representative 2D projection (bottom panel).

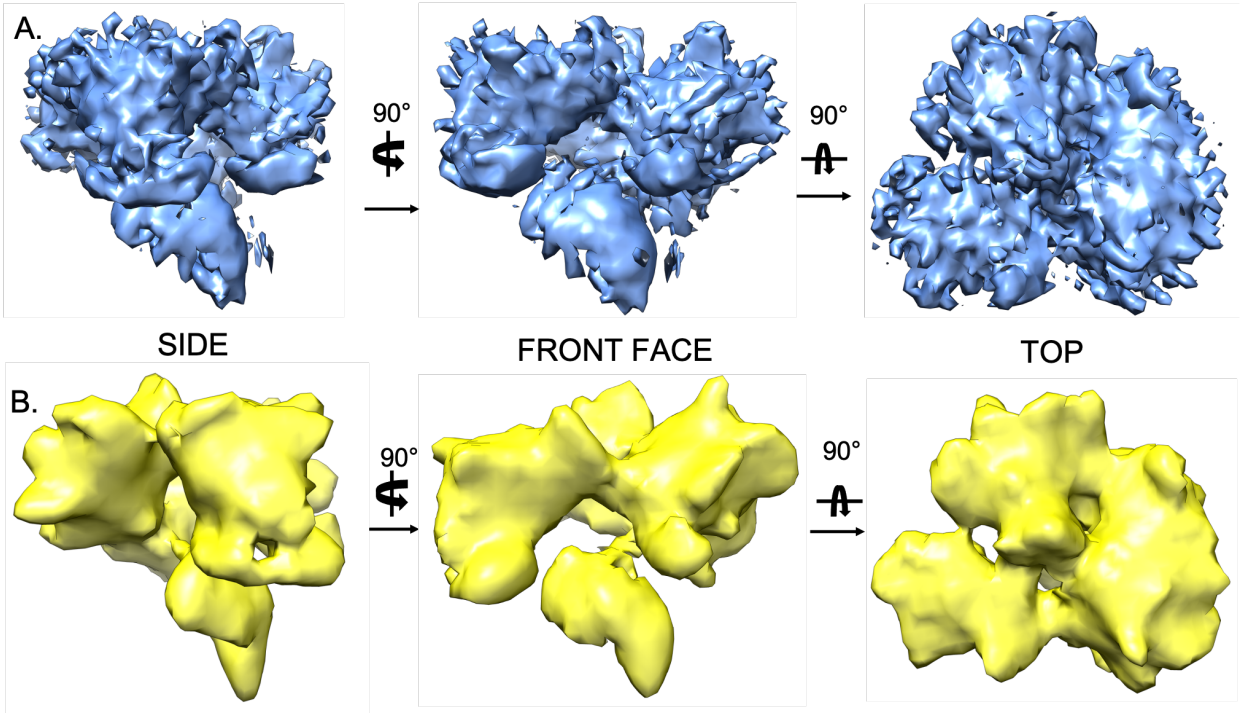


Figure 3.11. Human CBP•$(p53)_4$ Binary Complex (Conformation One) Structure Comparisons. Depiction of **(A)** the Initial Refinement map and (B) the Final Refinement map for the human CBP•$(p53)_4$ binary complex (conformation one) shown from three viewing angles (side, front, and top.)

The second arrangement of the human CBP-p53 complex was identified through a similar cryo-EM data processing strategy. Again, with no symmetry imposed, refinement and reconstruction of the second conformer from the selected particle projections showed an estimated overall global resolution of 11.0 Å resolution and a global resolution of 7.2 Å with further refinement (**Figure 3.12-3.14**). Interestingly, in stark contrast to the first structural state identified for the CBP-p53 sample, the second complex conformation took on an elongated and more open configuration with an entirely different overall shape. While the overall quadripartite character of the complex is conserved, it assumes a form resembling a J-shaped pitcher with tapering density from top to bottom **(Figure 3.15-Figure 3.17**). What appears to be the major lobe in the first conformation of the CBP-53 complex, forms the handle of the pitcher in this new open conformation, with densities of the base and right rim of the pitcher potentially, corresponding to the lobe 1 and lobe 2 of conformer one respectively. The top of the pitcher shows a donut-like density forming the central opening the pitcher. As with the first structural state, the second structural state of

the human cbp-p53 complex has pronounced crevices and protrusions present through the topology of the structure.

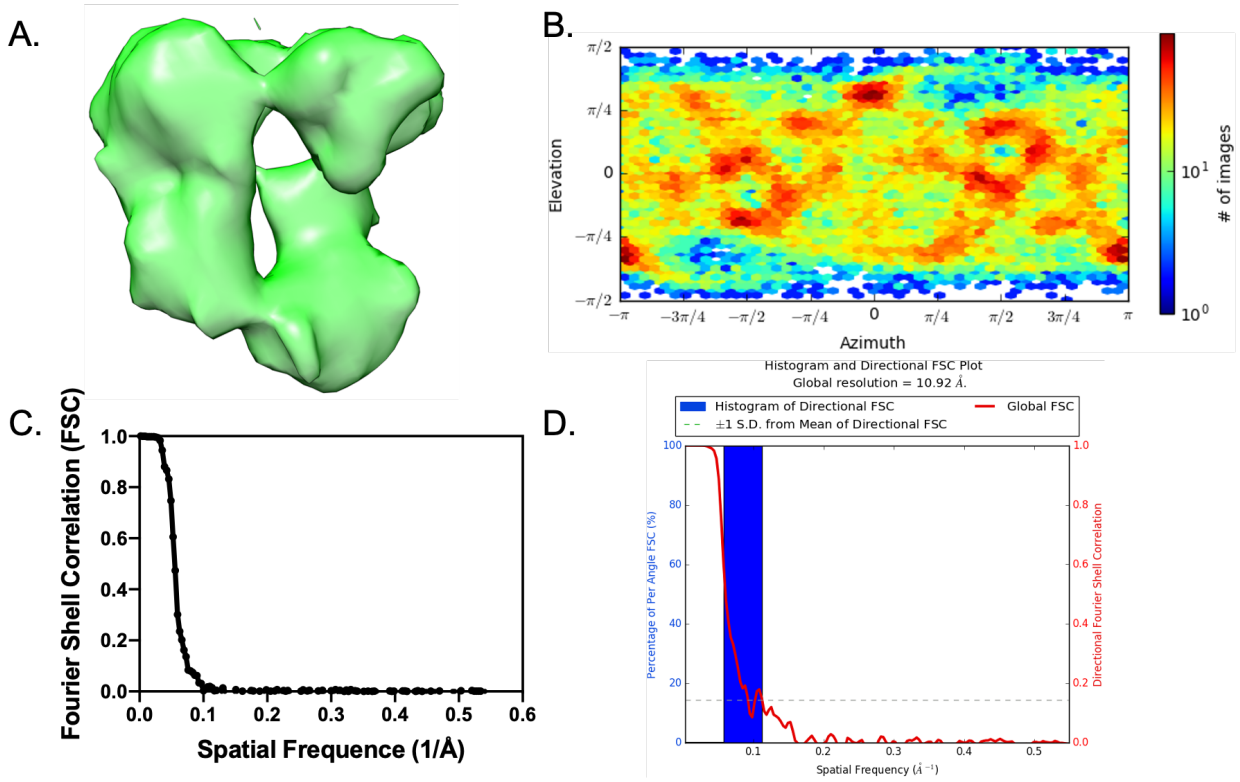


Figure 3.12 Cryo-EM Three-Dimensional (3D) Reconstruction of the human CBP•$(p53)_4$ binary complex (Conformation Two) and Validation of cryo-EM Reconstruction. (A) 3D reconstruction of the human CBP•$(p53)_4$ binary complex. (conformation two). The estimated resolution of the map is 11.0 Å. **(B)** Euler angle distribution of the human CBP•$(p53)_4$ binary complex (conformation two) highlighting some preferred orientations. **(C)** Gold Standard Fourier Shell Correlation (FSC) curve for conformation two of the human CBP•$(p53)_4$. The final resolution was determined using FSC criterion FSC=0.143 criterion. **(D)** Directional FSC histogram for the final 3D reconstruction of the complex.

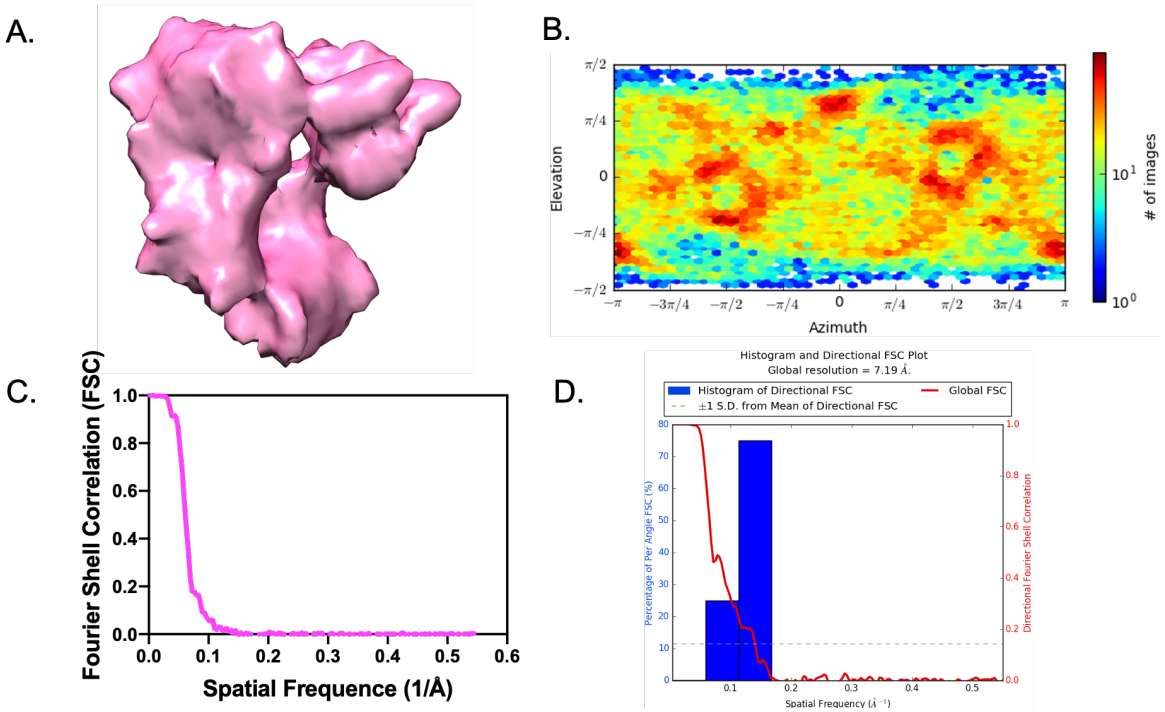


Figure 3.13 Cryo-EM Three-Dimensional Reconstruction of the Human CBP•$(p53)_4$ Binary Complex (Conformation Two) and Validation of cryo-EM Reconstruction After Further Refinement. (A) Three-dimensional reconstruction of the human CBP•$(p53)_4$ (conformation two) binary complex. The estimated resolution of the map is 7.2 Å. **(B)** Euler angle distribution of the human CBP•$(p53)_4$ (conformation two) binary complex highlighting some preferred orientations of the come. **(C)** Gold Standard Fourier Shell Correlation (FSC) curve for conformation two of the human CBP•$(p53)_4$. The final resolution was determined using FSC criterion FSC=0.143 criterion. **(D)** Directional FSC histogram for the final 3D reconstruction of the complex.

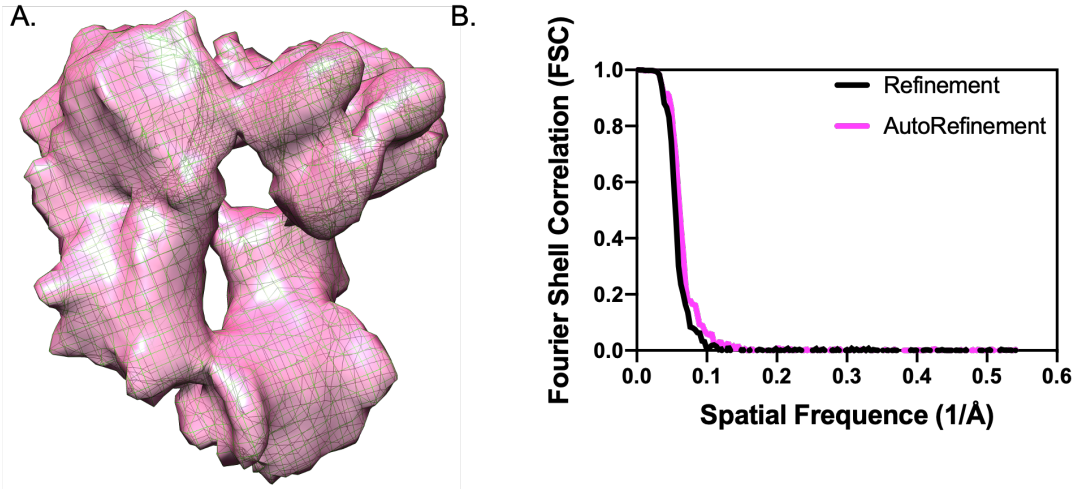


Figure 3.14. Human CBP•$(p53)_4$ Binary Complex (Conformation Two) Refinement Comparisons.
(A) The cryo-EM density map for the initially refined structure of the human CBP•$(p53)_4$ binary complex (green mesh) is superimposed on the final structure refinement (pink solid). **(B)** Overlayed initial refinement (black) and final refinement (pin) Gold Standard Fourier Shell Correlation (FSC) curves for conformation two of the human CBP•$(p53)_4$ complex.

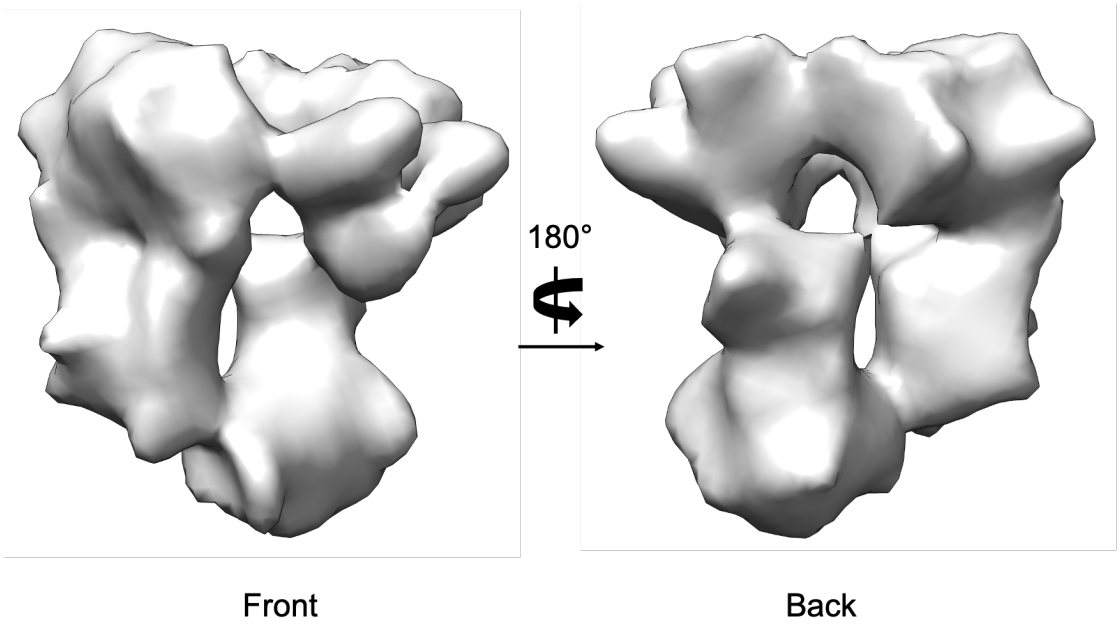


Figure 3.15. Overall Structure of The Human CBP•$(p53)_4$ Binary Complex (Conformation Two). The human CBP•$(p53)_4$ binary complex (conformation two) adopts the shape of a pitcher, tapered throughout its length.

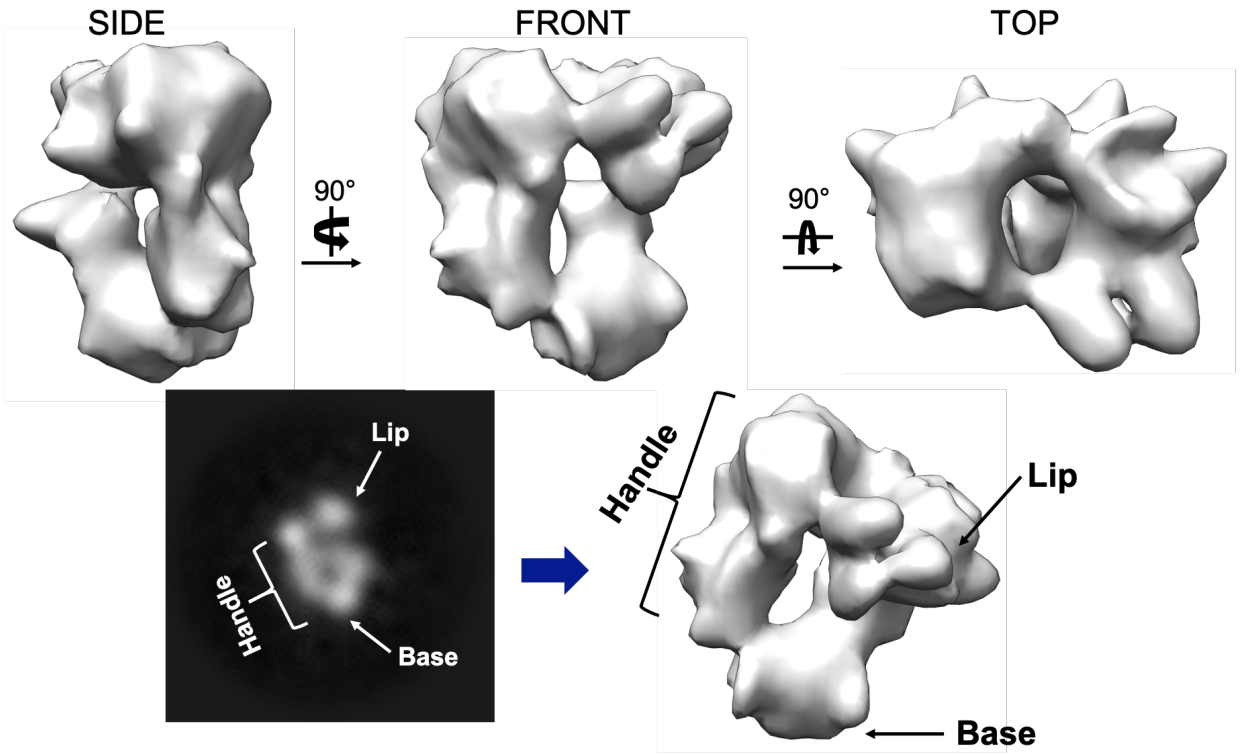


Figure 3.16. Structural Features of the Human CBP•$(p53)_4$ Binary Complex (Conformation Two). The overall architecture of human CBP•$(p53)_4$ binary complex (conformation two) has the shape of pitcher with a left protruding handle, narrow opening at the top, and small base. Depiction of the 3D reconstruction is shown from three viewing angles side, front, and top (upper panel) and is compared with a representative 2D projection (bottom panel).

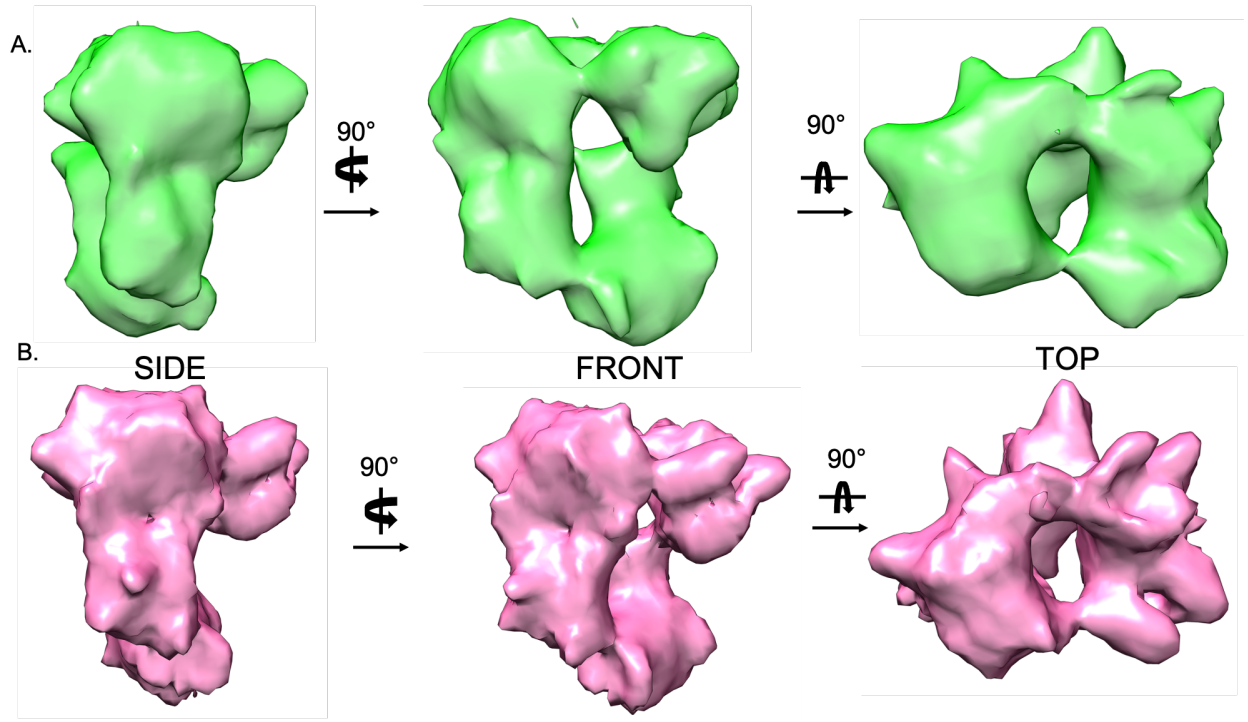


Figure 3.17. Human CBP•$(p53)_4$ Binary Complex (Conformation Two) Structure Comparisons. Depiction of **(A)** the Initial Refinement map and **(B)** the Final Refinement map for the human CBP•$(p53)_4$ binary complex (conformation two) shown from three viewing angles (side, front, and top).

Furthermore, the appearance of the CBP-p53 binary complex structures when compared to that of the full-length mouse p53 tetramer, suggest that one p53 tetramer is capable of interacting with each of the of four activator binding domains of the CBP as shown by the observable extra densities on the lobes of in the CBP-p53 binary complex **(Figure 3.18).** However, more studies focused on domain identification within the complex as discussed below will be required to validate this observation.

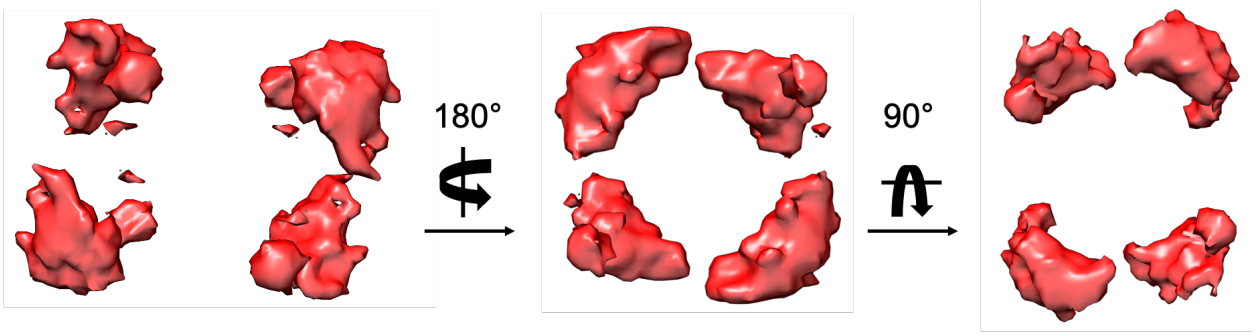


Figure 3.18 Three-Dimensional (3D) Reconstruction of the Full-length Murine Tumor Suppressor Protein p53. The tetrameric full-length murine p53, depicted from three viewing angels, has D2 symmetry and shows an overall shape that resembles a hollow skewed cube. The reported resolution is 26.0 Å.

For further comparisons of the structural heterogeneity present in the final structures, density difference mapping was carried out in addition to 3D variance analyses. For the density difference mapping the final higher resolution reconstruction of conformer two was normalized and aligned to that of conformer one, scaled, and subtracted to generate a difference map corresponding to extra densities presence. As shown in **Figure 3.19**, superimposition of the open configuration structure (conformer 2, blue mesh) with the closed binary complex (pink solid) highlight new displaced density elements presence, likely an outcome of large conformational heterogeneity within the overall complex itself. This suggests that the individual complexes themselves may adopt multiple conformations.

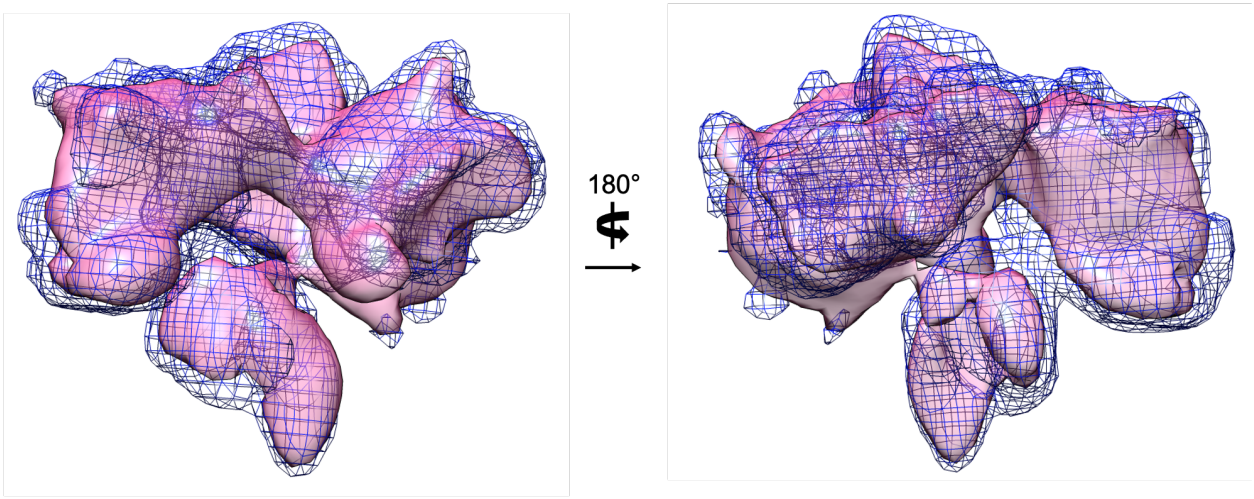


Figure 3.19. Conformation Comparisons of the Two Structural States of the Human CBP•$(p53)_4$ Complex. The cryo-EM structure of conformation one of the Human CBP•$(p53)_4$ superimposed to difference density map (blue mesh) calculated by subtracting the human CBP•$(p53)_4$ (conformation one, pink) from of the human CBP•$(p53)_4$ (conformation two). Conformational shifts are observed as indicated by the protruding densities of the difference map (blue mesh).

To further explore this concept, I used variance mapping in cryoSPARC to highlight regions of variance within each structure **(Figures 3.20-3.21)**. For each reconstruction the most variance seen can be localized to the larger lobes of the assemblies. **(Figure 3.24-3.25)**. Consistent with these observations, local resolution maps show lower overall resolutions for those regions that exhibit a higher degree of variability **(Figure 3.24-3.25)**.

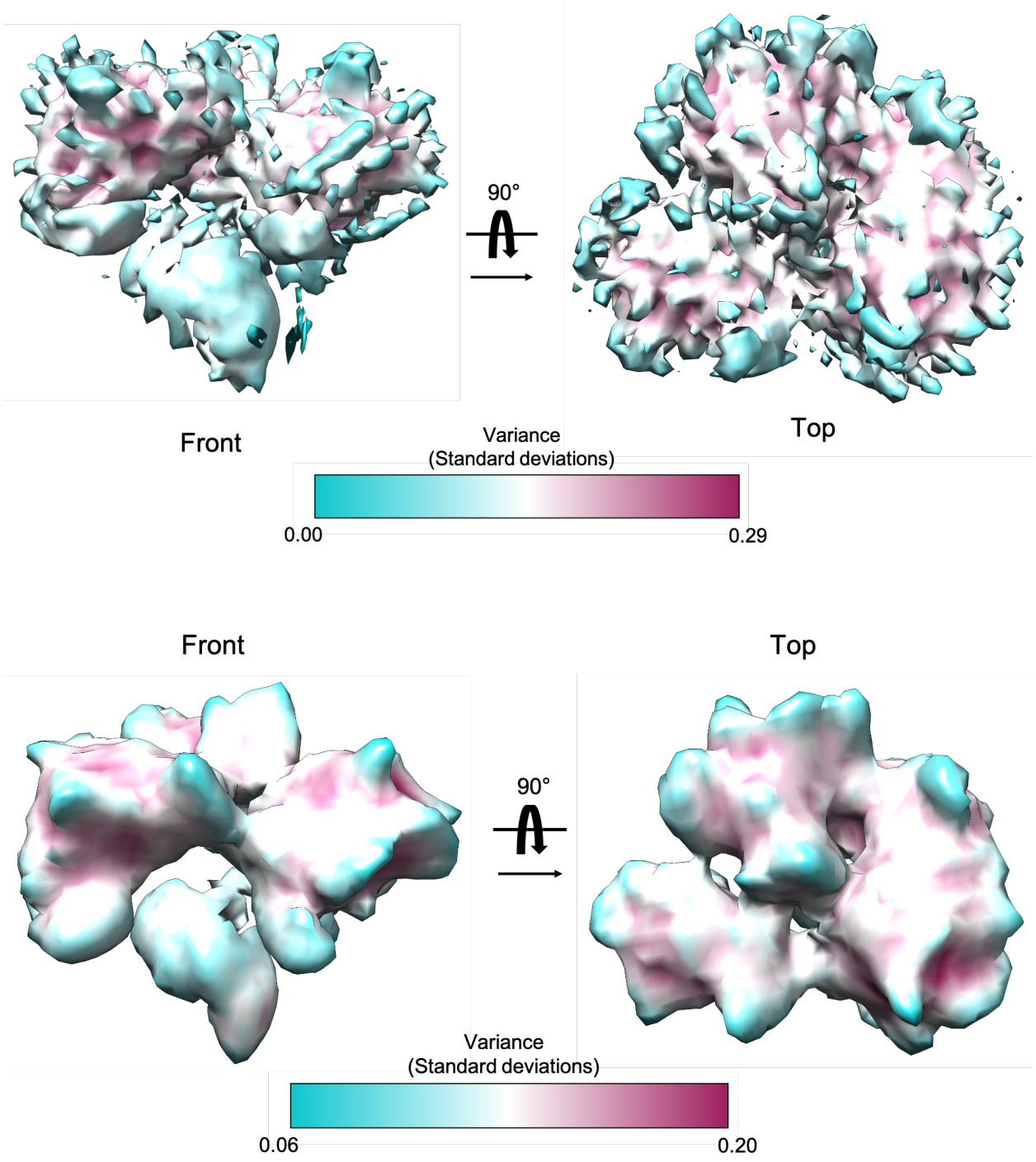


Figure 3.20 Three-Dimensional (3D) Variance Analysis of Conformation One of the Human CBP•$(p53)_4$ Binary Complex. Initial refinement (top panel) 3D statistical variance map calculated for the human CBP•$(p53)_4$ binary complex shows local flexibility. The local refinement with mask reduces the flexibility (bottom panel). Volumes are colored according to the variance from low (turquoise) to high (mulberry). Variance is given as standard deviation.

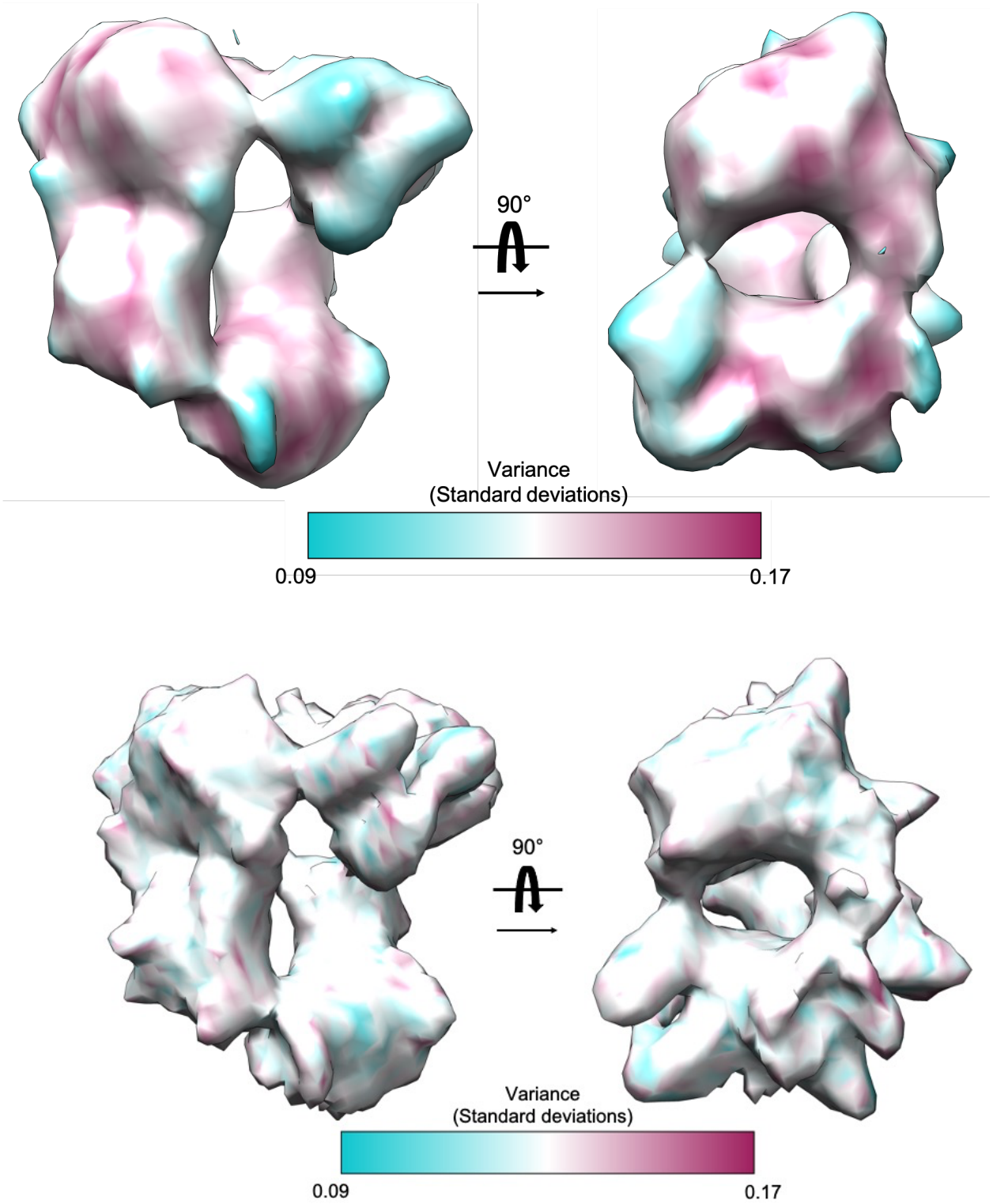


Figure 3.21 Three-Dimensional (3D) Variance Analysis of Conformation Two of the Human CBP•$(p53)_4$ Binary Complex. Initial refinement (top panel) 3D statistical variance map calculated for the human CBP•$(p53)_4$ binary complex shows local flexibility. The local refinement with mask reduces the flexibility (bottom panel). Volumes are colored according to the variance from low (turquoise) to high (mulberry). Variance is given as standard deviation.

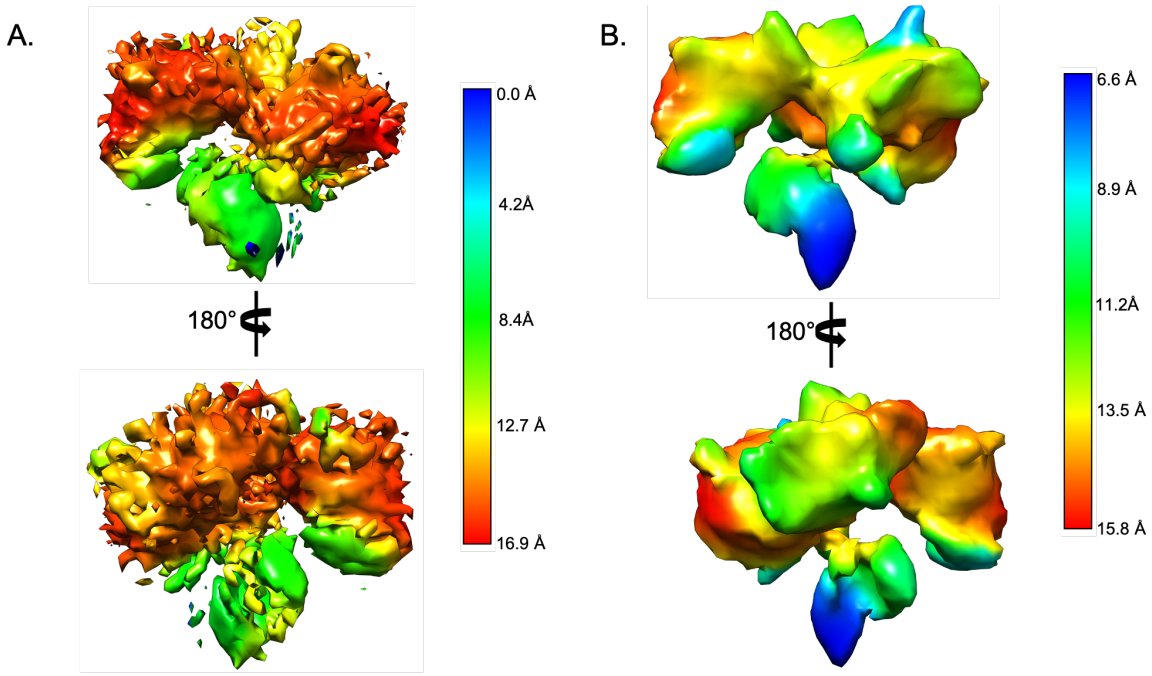


Figure 3.22 Local Resolution of Cryo-EM Map for the Human CBP•$(p53)_4$ Binary Complex (Confirmation One). Three-dimensional maps of **(A)** the human CBP•$(p53)_4$ complex (conformation one); initial refinement and **(B)** the human CBP•$(p53)_4$ complex (conformation one); local refinement. Density maps are colored according to local resolution estimations.

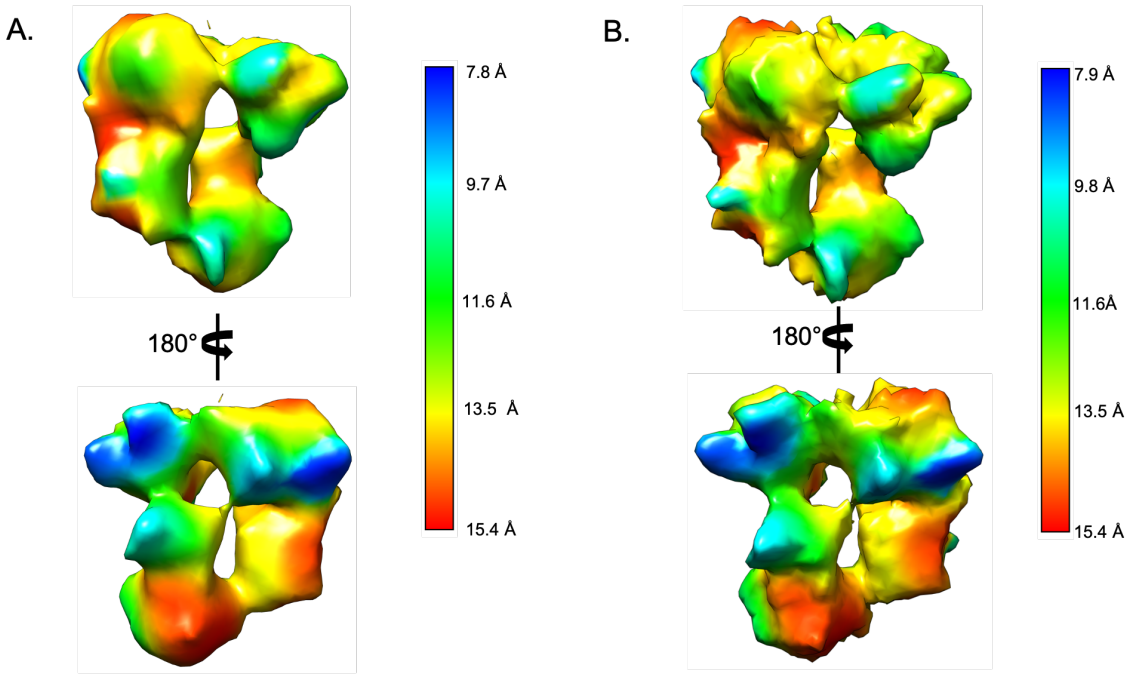


Figure 3.23 Local Resolution of cryo-EM Map for the Human CBP•$(p53)_4$ Binary Complex (Conformation Two). Three-dimensional maps of the (A) the human CBP•$(p53)_4$ complex (conformation one); initial refinement and (B) the human CBP•$(p53)_4$ complex (conformation two); final refinement. Density maps are colored according to local resolution estimations.

3.3 Conclusions

More than Just the Sum of the Parts

All together these findings provide the first full structural description on the intact coactivator-activator complex formed between CBP and p300. By employing extensive particle sorting and projection matching snapshots of two unique structural states of the human-CBP complex were captured, albeit at modest resolutions. The first structure looked much like a three-leaf clover whereas the second conformer in a more “open” configuration had the resemblance of a beer pitcher. The upper module of conformer one may serve as an anchor point for doubled stranded DNA to bind and dock to the top surface; the central cavity in the configuration of conformer 2, could easily accommodate double stranded DNA permitting binding and looping of the DNA throughout the length of the complex **(Figure 4.1)**. Both structural organizations allow for inter- and intra-domain interactions.

Of note, throughout the refinements and the heterogenous refinements (data not shown) of the human CBP-p53 complex, it was evident that the complex was structurally heterogenous and as a result some reconstructions contained regions of discontinuity, suggesting the presence of even more conformational states. Since CBP and co-activator complexes can interact with many different activators that have little to none structural homology, conformational flexibility would be advantageous for function. However, additional studies beyond the scope of this work to explore the full conformational landscape of this important transcription complex will be requisite inclusive of comprehensive studies with DNA. Moreover, while it would be of great interest to pursue these observations at higher resolution, first the identification and characterization the spatial arrangement of domains within this binary complex is perhaps the next most critical and logical step towards providing a more detailed description of the molecular recognition principles governing CBP and p53 complex formation.. Nonetheless these studies provide novel insights into how CBP and p300 might engage DNA and other binding partners and the current models proposed represent starting hypotheses to test going forward.

3.4 Experimental Methods

Protein Expression and Purification.

Protein expression was performed as previously described in chapter two**.** In brief Full-length CBP and p53 were co-expressed in SF9 cells. 48 hours post-infection cells were harvested and clarified by centrifugation. The CBP-p53 complex was then isolated by a three-step purification scheme inclusive of nickel Affinity, flag affinity, and size exclusion chromatography steps. The purified complex was then characterized by SDS-PAGE and visualized with negative stain electron microscopy to assess the suitability of the sample for cryo-EM studies. All steps were performed, cell harvesting through vitrification, were performed on a single day to limit degradation of the isolated complex.

*Cryo-EM Specimen Preparation and Data Acquisition**.***

3.5mL complex aliquots @ 0.3mg/mL were applied onto glow-discharged quantifoil grids, blotted with a Vitrobot (FEI Company) and flash frozen in liquid ethane. The grids were imaged at liquid nitrogen temperature on a FEI Talos Arctica electron microscope operating at 200 kV. Cryo-EM images were recorded with a Gatan K2 Summit direct electron detector at a nominal magnification of 45000X in counted mode, corresponding to a pixel size of 0.91A˚/pixel. A dose rate of 64.15 electrons/A°^2 /s and defocus values ranging within 1.5nm-3.0 nm were used. Total exposure of 8 s per image were dose-fractionated into 40 movie frames.

*Cryo-EM Data Processing Image Processing and 3D Reconstruction**.***

Micrograph movie stacks were first subjected to MotionCorr2 for whole-frame and local drift correction. Visual inspection was conducted to discard micrographs with visible astigmatism, large carbon areas, or ice contamination resulting in a working dataset of 2,412 cryo-EM images. CTFFIND4 was used for contrast transfer function (CTF) parameter determination. Approximately 2,000 particles were interactively picked and windowed in 324- pixel boxes with cryoSPARC v2.15. These particles were then subjected to reference-free 2D classification and averages were used to conduct automatic particle picking. The results of the automated picking were used in the final rounds of 2D averaging. Two resulting sets of particles with 49,547 particles and 63 classes and 46,222 particles and 49 classes were used to build ab initio reconstructions of the complex. Ab initio reconstructions with well-defined structural features were

subjected to iterative rounds of refinement. Resolution estimations for refinements were obtained with gold standard Fourier shell correlation using the 0.143 cut-off criterion. Depiction of molecular models were generated using UCSF Chimera.

3.5 References

1. Baldwin, P.R., et al., Big data in cryoEM: automated collection, processing and accessibility of EM data. Current Opinion in Microbiology, 2018. 43: p. 1-8.
2. Benjin, X. and L. Ling, Developments, applications, and prospects of cryo-electron microscopy. Protein Science, 2020. 29(4): p. 872-882.
3. Binshtein, E. and M.D. Ohi, Cryo-Electron Microscopy and the Amazing Race to Atomic Resolution. Biochemistry, 2015. 54(20): p. 3133-3141.
4. Bista, M., S.M. Freund, and A.R. Fersht, Domain-domain interactions in full-length p53 and a specific DNA complex probed by methyl NMR spectroscopy. Proceedings of the National Academy of Sciences, 2012. 109(39): p. 15752-15756.
5. Bose, D.A., et al., RNA Binding to CBP Stimulates Histone Acetylation and Transcription. Cell, 2017. 168(1-2): p. 135-149.e22.
6. Casañal, A., S. Shakeel, and L.A. Passmore, Interpretation of medium resolution cryoEM maps of multi-protein complexes. Current Opinion in Structural Biology, 2019. 58: p. 166-174.
7. Ciferri, C., et al., Molecular architecture of human polycomb repressive complex 2. eLife, 2012. 1.
8. Dutta, S., et al., Structure of a modular polyketide synthase. Nature, 2014. 510(7506): p. 512-517.
9. Fernandez-Leiro, R. and S.H.W. Scheres, A pipeline approach to single-particle processing inRELION. Acta Crystallographica Section D Structural Biology, 2017. 73(6): p. 496-502.
10. Joerger, A.C. and A.R. Fersht, Structural biology of the tumor suppressor p53. Annu Rev Biochem, 2008. 77: p. 557-82.
11. Joerger, A.C. and A.R. Fersht, The tumor suppressor p53: from structures to drug discovery. Cold Spring Harb Perspect Biol, 2010. 2(6): p. a000919.
12. King, L.A., R. Hitchman, and R.D. Possee, Recombinant Baculovirus Isolation, in Methods in Molecular Biology. 2016, Springer New York. p. 73-94.
13. Krois, A.S., H.J. Dyson, and P.E. Wright, Long-range regulation of p53 DNA binding by its intrinsically disordered N-terminal transactivation domain. Proceedings of the National Academy of Sciences, 2018. 115(48): p. E11302-E11310.
14. Krois, A.S., et al., Recognition of the disordered p53 transactivation domain by the transcriptional adapter zinc finger domains of CREB-binding protein. Proceedings of the National Academy of Sciences, 2016. 113(13): p. E1853-E1862.
15. Kussie, P.H., et al., Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain. Science, 1996. 274(5289): p. 948-953.
16. Larsen, K.P., et al., Architecture of an HIV-1 reverse transcriptase initiation complex. Nature, 2018. 557(7703): p. 118-122.
17. Mcsweeney, D.M., S.M. Mcsweeney, and Q. Liu, A self-supervised workflow for particle picking in cryo-EM. IUCrJ, 2020. 7(4): p. 719-727.
18. Melero, R., et al., Electron microscopy studies on the quaternary structure of p53 reveal different binding modes for p53 tetramers in complex with DNA. Proceedings of the National Academy of Sciences, 2011. 108(2): p. 557-562.
19. Melero, R., et al., Electron microscopy studies on the quaternary structure of p53 reveal different binding modes for p53 tetramers in complex with DNA. Proc Natl Acad Sci U S A, 2011. 108(2): p. 557-62.

20. Ohi, M.D., Structural and functional analyses of the spliceosome requires a multi-disciplinary approach. Methods, 2017. 125: p. 1-2.
21. Ohi, M., et al., Negative staining and image classification — powerful tools in modern electron microscopy. Biological Procedures Online, 2004. 6(1): p. 23-34.
22. Okorokov, A.L., et al., The structure of p53 tumour suppressor protein reveals the basis for its functional plasticity. The EMBO Journal, 2006. 25(21): p. 5191-5200.
23. Park, S.H., et al., Cryo-EM structure of the human MLL1 core complex bound to the nucleosome. Nature Communications, 2019. 10(1).
24. Peisley, A. and G. Skiniotis, 2D Projection Analysis of GPCR Complexes by Negative Stain Electron Microscopy, in Methods in Molecular Biology. 2015, Springer New York. p. 29-38.
25. Punjani, A., et al., cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nature Methods, 2017. 14(3): p. 290-296.
26. Punjani, A., H. Zhang, and D.J. Fleet, Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. Nature Methods, 2020. 17(12): p. 1214-1221.
27. Robinson, P.J., et al., Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. Cell, 2016. 166(6): p. 1411-1422 e16.
28. Rohou, A. and N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. Journal of Structural Biology, 2015. 192(2): p. 216-221.
29. Rout, M.P. and A. Sali, Principles for Integrative Structural Biology Studies. Cell, 2019. 177(6): p. 1384-1403.
30. Scheres, S.H.W., Processing of Structurally Heterogeneous Cryo-EM Data in RELION, in Methods in Enzymology. 2016, Elsevier. p. 125-157.
31. Schier, A.C. and D.J. Taatjes, Structure and mechanism of the RNA polymerase II transcription machinery. Genes Dev, 2020. 34(7-8): p. 465-488.
32. Skiniotis, G. and D.R. Southworth, Single-particle cryo-electron microscopy of macromolecular complexes. Microscopy (Oxf), 2016. 65(1): p. 9-22.
33. Stark, H. and A. Chari, Sample preparation of biological macromolecular assemblies for the determination of high-resolution structures by cryo-electron microscopy. Microscopy, 2016. 65(1): p. 23-34.
34. Stolt-Bergner, P., et al., Baculovirus-driven protein expression in insect cells: A benchmarking study. Journal of Structural Biology, 2018. 203(2): p. 71-80.
35. Takizawa, Y., et al., While the revolution will not be crystallized, biochemistry reigns supreme. Protein Science, 2017. 26(1): p. 69-81.
36. Veprintsev, D.B., et al., Core domain interactions in full-length p53 in solution. Proceedings of the National Academy of Sciences, 2006. 103(7): p. 2115-2119.
37. Weinberg, R.L., D.B. Veprintsev, and A.R. Fersht, Cooperative Binding of Tetrameric p53 to DNA. Journal of Molecular Biology, 2004. 341(5): p. 1145-1159.
38. Wells, M., et al., Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proceedings of the National Academy of Sciences, 2008. 105(15): p. 5762-576

CHAPTER FOUR

Conclusions and Future Directions from a Young, Black, and Gifted Scientist: Unifying Structural Differences and Collaborating for the GLOBAL Good

4.1 Introduction

Black Women and IDRS: Intelligently Designed and Forever Relevant

Conformationally dynamic proteins possess unique functional capabilities not readily achievable by more rigid proteins, making them of paramount importance to all biological processes. As such, eukaryotes have evolved a large and diverse set of flexible proteins that work together in a highly coordinated fashion to efficiently execute context-dependent cellular functions and processes. Gene transcription represents one such cellular process of life that is precisely performed through the use of dynamic molecular machines interconnected through an intricate web of protein-protein interactions. Coactivators make up one critical class of proteins involved in transcriptional activation whose structural plasticity make them perfectly suited to serve as the point of convergence for multiple signaling pathways. This structural plasticity confers them with the ability to undergo adaptive structural transitions to respond to a rapidly changing cellular milieu and set of incoming external cues. Importantly, this flexibility makes coactivators also highly adept at forming specific interactions with an enormously diverse set of macromolecules, including gene-specific transcription factors.

As a case in point, the master coactivator proteins CBP and p300 represent two malleable proteins that contain several conformationally dynamic domains. The conformationally dynamic domains of CBP and p300 are linked by intervening regions of disorder and have been shown to engage with hundreds of transcription activators. Several studies have suggested that conformational arrangements in the coactivators CBP and p300 and the complexes they form with activators are an essential component of mechanism. Conceivably, it can then be hypothesized that tight control over transcription is achieved through activator binding induced structural remodeling in coactivators. Structural determination of full-length transcription complexes, like the ones formed between CBP and p300 would represent an important step towards testing this hypothesis. Ironically, the very qualities that make CBP and p300 and their complexes exceptional in their ability to regulate transcription including their inherent flexibility, modular nature, transient interactions make them equally challenging to study structurally. As a result, a complete description of the architecture of CBP or p300 and how this architecture changes with activator binding has remained absent until now. Through the rigorous combination of biochemical characterization and single particle EM analyses on the full-length intact CBP-p53 transcription coactivator activator complex, the studies presented here represent the first structural description and analysis of the full, intact CBP-p53 transcriptional coactivator•activator complex and highlight important insights into the structural principles accounting for coactivator functions.

4.2 Summary of Dissertation Work and Overall Conclusions

Black Women and Science*: CHANGING* the Paradigm, and Contributing to Science and Society for a Lifetime

Structural studies on the CBP-p53 complex commenced with the development of an expression strategy and purification scheme to gently isolate sample for characterization. Following extensive efforts, co-expression of CBP and p53 to form the complex in insect cells was identified to be a viable option, overcoming the challenges that came with expression and purification of CBP and p53 individually, including a significant degree of heterogeneity and instability. Importantly, this expression strategy allowed for a preparative scale purification of a homogenous, relatively stable complex.

As assessed by SEC suitable for initial characterization by single particle negative stain EM and ultimately single particle cryo-EM.

After developing a reproducible expression and purification method to produce a sample suitable for single particle EM analysis CBP, the CBP•$(p53)_4$ binary complex and CBP•$(p53)_4$-DNA ternary complex were all examined by negative stain EM to gain further insight into the architecture of CBP and how this architecture is modulated by specific binding events. Importantly, comparisons of single particle classification and averaging on CBP•$(p53)_4$ with that of the ternary complex with DNA shows the overall shape of the complex becomes less extended and more compact upon DNA addition. These results provided initial support for the hypothesis that coactivators like CBP undergo structural rearrangement upon complexation with activators like p53. Nonetheless, because we observe that considerable flexibility remains in the complex, it may be that the structure can be further modulated by additional factors.

To gain higher resolution insights into coactivator-activator assembly and function, single particle cryo-EM was employed. Given the predicted conformation dynamics underpinning coactivator complex assembly and the expectation that the CBP•(p53)4 complex would likely exist in a number of conformations, a data processing pipeline inclusive of extensive project sorting and a multi-model refinement strategy was carefully designed and implemented. Unbiased reference 2D classifications were performed with the data set acquired and a number of distinct particle arrangements were observed, providing additional direct evident for the likelihood of several distinct states in the sample. Final refinements and reconstructions of the selected particle projections unveiled two conformational arrangements of the human CBP-p53 at modest resolutions. The overall shape of the conformers differ drastically. The first observed structural state of the human CBP•(p53)4 complex shows a more compact arrangement with three major and one minor lobe that orient in such a manner favoring the appearance of a three-leaf shamrock. The second conformer assumes a much more open and elongated state, assuming the form of a wide mouth pitcher that tapers through its length. Visual inspection of the architectural arrangement of the complex in both conformations when compared to the structure of the free p53 tetramer suggest that the p53 is interacting with each of the four-activator binding. With these structural arrangements, multiple modes for DNA binding

can be envisioned. In the first closed and more compact structural state, the density-rich upper module of the complex is well positioned to serve as a scaffold to capture DNA for initial binding, whereas in the second, more open arrangement of the complex DNA could potentially weave in and out of the complex**.** More detailed structural studies will be needed to support these conclusions.

All together the final structures presented within this dissertation define for the first time, the full architecture of the human CBP•$(p53)_4$ transcriptional coactivator-activator complex. Interestingly, but unsurprisingly the conformation of this complex is not static. The observations of flexibility in configuration of human CBP•$(p53)_4$ complex architecture the lend compelling support to outstanding hypotheses that suggest binding of p53 at one or more ABDs in CBP lead to large scale global structural reorganizations in the coactivator. Complementary, 3D variance analyses also demonstrate that both configurations display local variability, highlighting that intrinsic flexibility is likely a hallmark of coactivator function. Broadly speaking, structural transitions of the CBP-p53 complex would make permissible higher degrees of selectivity when forming additional interactions with other proteins, enhancing binding affinity of some interactions and decreasing binding affinity in others. Furthermore, conformation rearrangements could also have specific influences on DNA binding and the associated enzymatic activity of the HAT domain of CBP. In all proposed scenarios, the involvement of the other activators, general transcription factors, and other components of the transcriptional apparatus including RNAP II are likely to be required for full manifestation of consequential interactions and would be contingent upon specific cellular contexts. Collectively, the structural studies presented help to further define the principles of molecular recognition, assembly, and functional modulation governing the PPIs responsible for the vital biological process of gene expression. Critically, the observations from this investigation have significant functional implications for how coactivators perform their unique roles in transcription and point to a mechanistic model that greatly extends their regulation and regulatory potential.

4.3 Future Directions

Black Women: Breaking Through Barriers
“We’ve Only Come So Far. We Still Have a Long Way Go”

Future work on this system should include additional data processing, including further refinements and structure validation to ensure the accuracy of refined structural models presented. Localization of the individual domains of CBP and p53 is another much needed next step to establish the molecular arrangement and placement of each proteins. Protein domain assignment will greatly facilitate the interpretation of future and previously published biochemical and functional data on this system. Moreover, the definitive localization of individual domains of both CBP and p53 will serve as the basis to help answer lingering questions surrounding what interfaces are most important for assembly and molecular recognition in the context of full coactivator-activator complexes and how these interfaces are modulated.

Lastly more comprehensive functional and structural studies beyond the scope of this work will greatly add to in our understanding of how the coactivator conformational landscape changes with different variables. Additional functional and structural studies should include those with DNA, other co-factors, and small molecule ligands and should encompass mutational analyses and activity assays to augment current data. As the current study does not allow for the direct observation of any possible intermediate states of the complex that may play an important role in specificity nor does it account for the structural or functional impact imposed by post translational modifications, the application and integration of other biophysical methods should be of utmost consideration. Furthermore, it is important to remain aware that the lack of additional accessory proteins in the complex means that the structure is likely to be different than the native state. Therefore, structural studies inclusive of cryo-electron tomography are invaluable in this regard. Taken together, these additional studies are expected to shed an even more profound light on mechanistic determinants of function of coactivators like CBP and p300.